EFFECTS OF BIOMOLECULAR CROWDING ON PROTEIN STABILITY

Mohona Sarkar

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in the partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Chemistry.

Chapel Hill
2013

Approved by:

Gary J. Pielak

Dorothy A. Erie

Linda L. Spremulli

Ashutosh Tripathy

Terrence G. Oas

ABSTRACT

MOHONA SARKAR: Effects of Biomolecular Crowding on Protein Stability
(Under the direction of Gary J. Pielak)

The intracellular milieu is complex, heterogeneous and crowded— an environment vastly different from dilute, buffered solutions where most biophysical studies are performed. The cytoplasm excludes about a third of the volume available to macromolecules in dilute solution. This exclusion arises from the sum of two phenomena: steric repulsions and chemical interactions, often called hard and soft interactions, respectively. Most efforts to understand crowding have focused on steric repulsions.

Globular protein stability is the difference in free energy between the compact, biologically functional native state and the ensemble of less compact, nonfunctional denatured state. The hard-core repulsive component of crowding stabilizes globular proteins because the decrease in available volume favors compact species. The effect of soft interactions can be stabilizing or destabilizing. Soft repulsive interactions reinforce the stabilizing influence of hard-core repulsions. However, the equilibrium is shifted towards the denatured state in systems dominated by attractive non-specific interactions, because unfolding exposes more reactive surface.

In Chapter 1, I introduce the concepts of hard and soft interactions in more depth and discuss how they are expected to affect the equilibrium thermodynamic stability of globular proteins.

In Chapter 2, I describe experiments that test these concepts by using *Escherichia coli* cell lysates as the crowding agents, chymotrypsin inhibitor 2 (CI2) as the test protein and NMR-detected amide-proton exchange to measure stability. The lysate destabilizes CI2, and the destabilization increases with increasing lysate concentration. This observation shows that the cytoplasm interacts favorably, but non-specifically, with CI2, and these interactions overcome the stabilizing hard-core repulsions. In fact, the effects of the lysate are even stronger than those of homogeneous protein crowders, reinforcing the biological significance of weak, non-specific interactions.

In Chapter 3, I test the idea that the net charge on the crowding proteins affects stability. To accomplish this goal, I isolated the anionic proteins from the lysate and used them as the crowding agent. CI2 is an anion under the chosen conditions, and, therefore, I expected the net repulsive interactions between CI2 and the crowders to increase the stability of CI2. Instead, the refined lysate also resulted in destabilization. Thus, even the anionic proteins, which have the same net charge as CI2, significantly interact with CI2's backbone non-specifically to overcome the stabilizing effect of steric repulsion.

My *in vitro* studies show that weak chemical interactions play key roles in the cytosol. It will even be more difficult to identify soft interactions in living cells where reductionist approaches are difficult or impossible to apply. Nevertheless, endeavors aimed at quantifying soft interactions are essential for producing a physiologically relevant understanding of biophysics. Once the details of soft interactions are known, it should be possible to tune them so as to obtain bespoke behavior in test tubes and in cells. In summary, despite their weak and non-specific behavior, biologists of all types need to keep these

interactions in mind when designing experiments to correlate *in vitro* studies with *in vivo* behavior.

To my parents, Dipankar Sarkar and Phuljhuri Sarkar, for instilling the scientific aptitude in me from childhood.

ACKNOWLEDGEMENTS

The last five years in graduate school would not have been such a wonderful journey in absence of the following people. First and foremost, I would like to thank my PhD advisor, Gary. I met him while I was grading for his CHEM 101 class and decided to rotate in his lab, because I liked his sense of humor and of course, the Amante's pizza he provided for the graders. While rotating in his lab, I really appreciated his patience in teaching me biophysics. He explained complex problems of biology in such a fundamental way, that it encouraged me to join his lab and pursue protein biophysics, even though I had little experience with protein chemistry and NMR. His door was always open for me for all types of silly questions and I thank him for always being there to help me overcome all the hurdles I faced in the last 5 years. He's indeed my superman! I am grateful to him and his wife, Elizabeth for inviting me to their home during Thanksgiving and Christmas. I am lucky to have an advisor like him. Gary and Elizabeth helped me in making Chapel Hill home for last five years. I will treasure his guidance and teachings in my journey in the path of science henceforth.

Secondly, I would like to thank all my friends and labmates, especially Andrew, Yaqiang, Austin and Joe, with whom I have worked closely on many of the projects. Andrew and Yaqiang trained me in NMR and biochemistry techniques. All those arguments with Austin helped improve my understanding of science. Moreover, my friends, Will, Austin, Jillian, Gauri, Deepak and Rachel have always been there to patiently hear my problems. An

informal discussion with Deepak over coffee resulted in the idea of charge dependent crowding.

I would also like to take this opportunity to thank my committee members, Drs. Dorothy Erie, Linda Spremulli, Ashutosh Tripathy and Terrence Oas, for their helpful insights and advice. I would like to thank Drs. Greg Young, Marc ter Horst and David Smalley for helping me with NMR and mass spectrometry.

Last, but not least, I am immensely grateful to my parents who have sacrificed a lot in their lives to help me grow as a responsible person and a scientist. My father always encouraged to take my failures in stride and to question my successes, thereby instilling scientific skepticism in me. My mother always believed in me to overcome all my problems, giving me the confidence to move forward during trying times. As a student of science, I can’t help but thank the wonderful biochemistry of caffeine and adrenaline, which helped me power through all those sleepless nights pondering over infinite unresolved tasks.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

Å	angstrom
°C	degree Celcius
$\Delta G_{op}^{o'}$	standard free energy of opening
$\Delta G_{D}^{o'}$	standard free energy of denaturation
$\Delta H_{D}^{o'}$	standard enthalpy of denaturation
$\Delta S_{D}^{o'}$	standard entropy of denaturation
µL	microliter
µM	micromolar
B_{23}	second virial coefficient
BSA	bovine serum albumin
CI2	chymotrypsin inhibitor 2
CRABP	cellular retinoic acid binding protein
D	ensemble of denatured states of protein
DEAE	diethylaminoethanol
DNA	deoxy ribonucleic acid
f	Mayer f-function
g	gram
x g	times gravity
h	hour
HSQC	heteronuclear single quantum coherence
In vitro	Latin for experiments performed in a test tube
In vivo	Latin for experiments performed in living organisms

IPTG	isopropyl β-D-1-thiogalactopyranoside
k_{cl}	rate of closing
kDa	kiloDalton
k_{int}	intrinsic rate of amide exchange
k_{op}	rate of opening
k_{obs}	observed rate of amide exchange
L	liter
LB	Luria broth
M	molar
mg	milligram
MHz	Megahertz
min	minute
mM	millimolar
N	native state of protein
NaCl	sodium chloride
nM	nanomolar
NMR	nuclear magnetic resonance
NOESY	nuclear Overhauser enhancement sprectroscopy
OD_{600}	optical density at 600 nanometers
PEI	polyethylenimine
p*I*	isoelectric point
PVP	polyvinylpyrrolidone
RNA	ribonucleic acid

s	second
SASA	solvent accessible surface area
SDev	standard deviation of the mean
SEM	standard error of the mean
SOLEXSY	solvent exchange spectroscopy
SPHERE	server program for hydrogen exchange rate estimation
vol	volume
v/v	volume/volume
w/v	weight/volume

CHAPTER 1: SOFT INTERACTIONS AND CROWDING

The material in this chapter is adapted from:

Sarkar M, Li C, Pielak GJ. Soft interactions and crowding. *Biophysical Reviews* 5:187-194 (2013).

1.1 Introduction:

For ideal solutions the measured concentrations of species exactly match their chemical behavior (*i.e.*, activity coefficients are unity), the enthalpy of solution is zero and the interactions between any two species are identical. Under these conditions, only water-solute interactions are important and all solutes are invisible to each other. The cellular environment deviates markedly from ideality. It is heterogeneous and crowded by many different macromolecules. For instance, macromolecules in the *Escherichia coli* cytoplasm can reach concentrations of 300 to 400 g/L and occupy 30 - 40% of the volume (1). In cells solutes not only interact with water, but also with the other cosolutes. Moreover, the interactions are neither chemically nor spatially similar. For example, any particular cellular protein can interact with other proteins, nucleic acids and other biomolecules. Capturing the effects of this non-ideality demands that biological macromolecules be studied in cell-like environments.

Cellular interiors are generally mimicked by using various macromolecules as "crowders". To understand the chemical nature of crowding effects, small cosolutes such as osmolytes are also often used. To facilitate observing the effect felt by one particular

macromolecule in the crowded environment, we introduce the concept of the test molecule, a species whose concentration is insignificant compared to the total concentration of macromolecules. Thus, test molecules rarely interact with each other. To detect the test molecule in a sea of the other macromolecules it must possess a unique probe, such as a fluorescent tag or isotopic enrichment.

Zimmerman and Trach (1) showed that cytoplasmic conditions change the activity coefficients of test molecules. Thus, the equilibrium thermodynamic behavior of these molecules is expected to differ in cells compared to dilute solutions.

Here, we are mostly interested in the equilibrium stability of globular proteins with two states (2): the efficiently-packed (3), biologically-active native state (N) and the inactive, denatured state (D). D comprises a large ensemble of conformations of the disordered protein (4), whereas N comprises a much smaller and more compact ensemble centered on the folded conformation. In other words, D is a thermodynamic state, while N is both a thermodynamic state and well-defined structural state. Protein stability is defined as $\Delta G_D^{o'}$, the modified standard state Gibb's free energy of D minus N. The stability at a given temperature is governed by the entropy and the enthalpy of each state, $\Delta G_D^{o'} = \Delta H_D^{o'} - T\Delta S_D^{o'}$. N possesses the lowest free energy, but D is entropically favored because it is less ordered than N (2). Increasing $\Delta H_D^{o'}$ and/or decreasing $\Delta S_D^{o'}$ increases protein stability by increasing $\Delta G_D^{o'}$.

Most mesophilic globular proteins are marginally stable in dilute solution near neutral pH at room temperature, possessing $\Delta G_D^{o'}$ values of 5 to 15 kcal/mol (5). The stability arises from the difference between two rather large (~10^2 kcal/mol) numbers, $\Delta H_D^{o'}$ and $T\Delta S_D^{o'}$. The cellular environment can affect stability by altering either or both of these terms. Part of the entropic contribution under crowded conditions arises from steric hard-core repulsions

between the crowding molecules and the test protein. As pointed out by Minton (6) in his ground-breaking work, these steric interactions are predicted to increase stability because N is more compact than D. Until recently most efforts to understand crowding effects have focused on this entropic component.

Enthalpic contributions are more subtle because they depend on chemical interactions between the crowder and either or both D and N. Attractive interactions with D and non-specific attractive interactions with the protein in general (*e.g.*, urea induced denaturation) lead to destabilization (7). On the other hand, attractive interactions with N (*e.g.*, ligand binding) tend to be stabilizing. Stabilization can also arise when the free energy of transferring a peptide bond from a dilute aqueous solution to an aqueous solution containing the cosolute is unfavorable (8). This preferential hydration of N is stabilizing because unfolding exposes more backbone to cosolutes (9). In summary, unlike hard-core repulsions, which are always stabilizing, non-specific interactions can be stabilizing or destabilizing.

1.2 Excluded volume

The entropic and the enthalpic contributions to protein stability can be approximately dissected into hard- and soft- (also called chemical-) interactions, respectively. The interplay of hard- and soft- interactions determines the excluded volume, a useful concept for understanding crowding effects.

The excluded volume, v, equals the negative volume integral of the Mayer f-function (10).

$$v = -\int \left(\exp\left[-\frac{U(r)}{kT}\right] - 1\right) d^3r \qquad \text{(i)}$$

U(r) is the interaction energy, which depends on the distance r between the particles. At small values of r, the interaction is highly repulsive because of the difficulty in interpenetrating the electron shells of the two species. The interaction is often attractive at larger values of r, but eventually decays to zero.

A pictorial representation is shown in Figure 1. The total shaded area is -1 times the excluded volume. With only hard-core repulsions, the excluded volume is just the volume that one crowder excludes from the test protein. If both particles are spheres, the hard-core excluded volume equals the co-volume; the volume of a sphere whose radius is the sum of the crowder and protein radii (11). Excluded volume arising from hard-core repulsion also increases with crowder size and concentration (6).

Although, we have focused on the simplest system, where both crowder and test proteins are hard spheres, even this type of excluded volume depends on both size and shape. This hard-particle excluded volume can be approximated by using two approximations: scaled particle theory and lattice theory. Scaled particle theory allows assessment of non-ideality for any hard convex particles in the midst of similarly shaped but differently sized hard particles (12, 13). Lattice theory, on the other hand, approximates the non-ideality of a hard particle of any shape in the midst of hard rectangular parallelepipeds (6).

1.3 Soft interactions and excluded volume

Soft interactions can add to or subtract from the hard-core excluded volume. The idea that repulsive chemical interactions increase the excluded volume is easy to grasp. For instance, if the crowder and test protein have the same charge, the resulting repulsion

increases the area below the x-axis in Figure 1. That is, repulsive interactions make the crowder appear larger, increasing the excluded volume.

The idea that attractive, non-specific interactions decrease excluded volume is less intuitive because it is difficult to picture how a particle can get smaller. Although its size does not change, the attractions increase the shaded area above the axis. The net result can be, as shown in Figure 1, that attractive interactions balance the repulsive interactions leading to a small –or even negative– excluded volume (14). In summary, soft interactions can be attractive or repulsive, either diminishing or enhancing, respectively, the excluded volume.

1.4 Excluded volume, protein stability and Le Chatelier's principle

The relationship between protein stability and the hard-core component of crowding is straightforward. Assuming that N and D are hard spheres, the fact that the radius of D is in some sense larger than that of N, means the volume excluded to D by crowding is larger than that excluded to N. Le Chatelier's principle tells us that the species taking up the least volume will be favored under crowded conditions, which means the protein is stabilized. This stabilizing effect is completely entropic because the model involves only the arrangement of molecules, not chemical interactions between them. Minton (6) was the first to fully describe these ideas in terms of biology. Later, Zhou et al. (15) also considered the equilibrium effects of crowding and confinement on biomolecular association, ligand binding and protein stability with emphasis on steric, hard-core repulsions. They concluded with remarks encouraging the incorporation of non-specific chemical interactions. Minton (16) also mentions the effects of hard and soft interactions on protein stability in his review on crowding and protein-protein interactions.

Non-specific, soft, attractive interactions between the crowder and the test protein tend to cancel this stabilizing hard-core effect.[1] Urea provides an example of a small molecule that acts non-specifically. This cosolute destabilizes proteins because it interacts favorably with the protein backbone. This effect is stronger for D than it is for N because D offers up more backbone for interaction (7, 17). Applying Le Chatelier's principle leads to the conclusion that urea favors the denatured state because D maximizes the number of attractive, non-specific interactions. These attractive interactions overcome the hard-core repulsion, and the protein unfolds. In summary, the net effect of crowding depends on the winner of a battle between attractive and repulsion interactions.

1.5 Measuring excluded volume

Assessing hard- and soft- interactions via their effect on thermodynamic non-ideality is one way of predicting the behavior of a system. The non-ideality is defined by the virial expansion of the chemical potential and the molar activity coefficient involving the solute-solute, solute-cosolute and cosolute-cosolute. In this expansion, each component is identified by a subscript: 1 is the solvent, 2 is the test protein, 3, *etc.* are cosolutes (18). $B_{23,}$ B_{223}, B_{233} *etc*., are cross-virial terms between the protein and the cosolutes. The latter two terms and others like them are unimportant in the studies discussed here because, by definition, the concentration of the test protein is small. The key term is 'B_{23}', the second virial co-efficient, which measures the protein-crowder interaction. When no chemical interactions are present and both species are spheres, B_{23} is the excluded volume as represented by the co-volume (11, 19).

[1] Non-specific attractive interactions are fundamentally different from ligand binding, because ligands bind only to N. Application of Le Chatelier's principle leads to the conclusion that ligand binding stabilizes proteins by favoring N.

$$B_{23} = (4/3)\pi N(r_2 + r_3)^3 \quad \text{(ii)}$$

where N is Avogadro's number. Thus, B_{23} measures the excluded volume as defined in equation (i), and its assessment can identify the source of crowding effects. Note that soft effects can be "fudged" by adjusting the size of the crowder in equation (2).

With the exception of Minton (20) on using osmotic pressure to define soft interactions involving Bovine Serum Album (BSA, 68 kDa), much of the work in this area has focused on the small molecule cosolutes. The Record group (21) used vapor pressure osmometry with the test protein BSA and six small cosolutes to obtain values of B_{23}. Their work focused on correlations between protein-cosolute interactions and changes in water-accessible surface area. Similar studies probed the interactions between the test protein cytochrome *c* and small cosolutes (22) using sedimentation equilibrium analytical ultracentrifugation. Both studies showed that B_{23} decreases in systems where the small molecules interact favorably with the test protein and increases in systems when there is enhanced repulsion. A more detailed assessment of second virial coefficients for small osmolytes is given by Davis-Searles *et al.* (23).

1.6 Crowding and assembly

Turning to the effects of macromolecular crowding on reactions, the self-association of the bacterial cell division protein FtsZ was studied by Rivas *et al.* (24). Using cyanomethemoglobin and BSA as crowders and non-ideal tracer sedimentation equilibrium ultracentrifugation for detection, they observed that crowding increases the weight average degree of the association of FtsZ. They concluded that pure hard-core excluded volume promotes the self-assembly. Hirota, *et al.* (25) examined the conversion of a linear

cytochrome *c* trimer to the smaller cyclic trimer in a system crowded with polyethylene glycol, and observed that crowding shifted the equilibrium toward the cyclic trimer. Jiao *et al.* (26) studied the effects of three synthetic polymers on the association of catalase and superoxide dismutase. They found that steric repulsions dominate at high temperatures, but protein-polymer non-specific interactions dominate at low temperatures.

1.7 Soft effects and the challenge of in-cell protein NMR

The ^{15}N-^{1}H Heteronuclear Single Quantum Coherence (HSCQ) experiment yields a protein's fingerprint (27). The covalent bond between the backbone amide ^{15}N and its proton gives rise to a feature, called a crosspeak, at the chemical shift coordinates of the two nuclei. For this reason the experiment is a powerful tool for obtaining residue-level information from ^{15}N-enriched proteins. Resonance broadening with the resultant loss of resolution and sensitivity, however, makes HSQC-based in-cell NMR a challenge for the study for globular proteins (28-32).

To obtain a high resolution HSQC spectrum the test protein must have considerable rotational motion (27). Increasing the molecular weight of the protein or the solution viscosity decreases its rotational motion, resulting in poor quality spectra. The cellular environment seems to increase the molecular weight of test proteins because poor or non-existent HSQC spectra are obtained for proteins that otherwise give outstanding dilute solution spectra (33). The apparent increase in molecular weight could be the result of increased viscosity in cells and/or the presence of numerous weak, non-specific attractive interactions between the test protein and other cellular macromolecules.

1.8 Soft interactions are important in cells

NMR studies of protein diffusion and dynamics can help differentiate weak non-specific interactions from increased viscosity under crowded conditions. Li and Pielak (34) used an ^{15}N-relaxation experiment to identify the presence of attractive non-specific interactions involving the test protein Chymotrypsin Inhibitor 2 (CI2, 7.4 kDa) with BSA as the crowder. These interactions are also expected to affect the diffusion. Wang, Li and Pielak (35) used ^{15}N relaxation to observe deviations in CI2 rotational diffusion under crowded conditions. These authors crowded the system with glycerol, polyvinylpyrrolidone, Ficoll, lysozyme, ovalbumin, BSA and *E. coli* lysates. Rotational diffusion was impeded more in protein crowders than in synthetic polymers. These studies of nuclear relaxation showed that protein crowders and lysates interact favorably with CI2. In summary, attractive non-specific interactions increase the apparent molecular weight of globular test proteins in cells, impeding their rotational diffusion, which leads to poor quality, or non-existent in-cell HSQC spectra.

Crowley, Chow and Papkovskaia (32) studied the interactions of GB1 (6.2 kDa, pI 4.5) and cytochrome *c* (11.5 kDa, pI 9.6) with cytosolic *E. coli* proteins by using size exclusion chromatography of cell lysates. GB1, which gives a good in-cell HSQC spectrum, elutes at a volume commensurate with its calculated molecular weight, showing that it does not interact strongly with cytosolic proteins. On the other hand, the highly positively charged protein cytochrome *c,* which does not give an in-cell HSQC spectrum, elutes at lower volume, implying stronger interactions. At least some of the effect is caused by charge-charge interactions because increasing the ionic strength or introducing charge-inverting mutations increased the elution volume of cytochrome *c*.

Interestingly, HSQC spectra of disordered proteins such as FlgM and α-synuclein are easier to observe in cells (29, 36). This increased detectability probably arises because disordered proteins possess much more internal motion than do globular proteins, and this internal motion is less affected by attractive interactions with other cellular macromolecules.

To determine why globular proteins give such poor in-cell HSQC spectra, the Gierasch group studied three small globular proteins, GB1 (6.2 kDa), NmerA (6.9 kDa) and ubiquitin (8.7 kDa) (31). They observed good quality in-cell spectrum of monomeric GB1, but a poor quality spectrum of NmerA and no signal from ubiquitin. Their dimeric GB1 construct, although twice the size of NmerA, produced a reasonably high quality spectrum. They deduced that if viscosity alone caused the broadening, folded proteins smaller than 13 kDa would give good quality in-cell spectra, which was not the case. They then set out to identify the interactions that result in poor or useless HSQC spectra.

NmerA and ubiquitin have the same number of surface hydrophobic residues, but they are localized in ubiquitin and dispersed in NmerA. Decreasing the surface hydrophobicity of ubiquitin and altering its distribution by changing three valine residues to alanines, yielded sharper spectra in lysates, indicating that the hydrophobic effect contributes to soft interactions in cells. Charge-charge interactions also play a role. The environment inside *E. coli* is anionic (37). GB1 has a negative charge at physiological pH whereas NmerA and ubiquitin have insignificant net charge, making the latter two more 'sticky'. As discussed above, Crowley et al. (32) have also pointed out the importance of charge.

In summary, the absence of signals in in-cell NMR experiments can be explained by weak attractive interactions, including charge-charge and hydrophobic interactions. Next, we

turn to the effects of crowding on protein folding in cells and on the equilibrium thermodynamics of globular protein stability.

1.9 Protein folding in the cell

Schlesinger, Wang, Tadeo, Millet and Pielak (38) tested the ability of the *E. coli* cytoplasm to fold a globular protein in cells. Wild type Protein L (ProtL, 7.0 kDa) and a disordered variant were studied. As expected, the wild type protein does not generate an in-cell HSQC spectrum. On the other hand, its Kx7E variant is disordered and, again as expected, yields a reasonable in-cell spectrum. Dilute solution studies showed that the variant required only 0.4 - 1.0 kcal/mol to fold (39). If macromolecular crowding effects in cells were predominately the result of hard-core repulsion, then the cellular environment would be expected to provide sufficient free energy to fold the protein. The protein, however, remained unfolded in cells. This observation suggests that soft interactions can overcome the stabilizing effects of hard-core repulsions.

1.10 *In vitro* effects on stability: crowding by synthetic polymers

With opposing phenomena playing key roles, it might be possible to predict the net effect of crowding given the details of the system. Hard-core repulsions will increase $\Delta G_{den}^{o'}$ by shifting the equilibrium towards N, because N is more compact. This phenomenon has been studied theoretically and experimentally. Theoretical models have been successful in mimicking the hard-core repulsion effect (e.g., 15, 24, 40, 41), and many 'wet' experiments show that crowding by synthetic polymers stabilizes proteins (e.g., 42, 43-47). These observations are consistent with theoretical expectations in as much as the stabilizing hard-

core repulsions seem to dominate, although significant attractive protein-crowder interactions have been noted for polyethylene glycol (25, 48, 49) and the monomer of polyvinylpyrrolidone (42).

We close this section with three *caveats*. First, synthetic polymers probably cannot be approximated as spheres and are certainly not impenetrable (50) making application of simple ideas such as co-volume problematic. An early simulation of excluded volume effects for polymer-protein interactions was performed assuming that polymers are flexible chains obeying Gaussian statistics (51). This assumption gave a more realistic picture. Second, crowding effects can be more complicated if the test protein is not spherical; Homouz *et al.* (52) combined simulation with wet experiments to show that macromolecular crowding affects both the shape and stability of an aspherical test protein, VlsE. Third, the role of preferential hydration (8) in the observed stabilization remains unclear.

1.11 *In vitro* effects on stability: crowding by proteins

Mikos *et al.* (53) assessed CI2 stability in solutions crowded by BSA and lysozyme. In contrast to the stabilization observed with synthetic polymers, the protein crowders destabilized CI2. This observation suggests that the proteins interact favorably and non-specifically with the backbone of CI2 so as to overcome the stabilizing effect of hard-core repulsions. These authors also showed that a part of these weak non-specific interactions originate from charge-charge interactions, consistent with the in-cell experiments discussed above (31, 32). Molecular dynamic simulations showed that the backbone fluctuations of CI2 increase in lysozyme, which correlates with its destabilizing effect (54). Harada et al. (55) used molecular dynamics to show that protein crowders reduce the dielectric constant of the

solution. This decrease would diminish the hydrophobic effect while enhancing hydrogen bond strength. In summary, studies with protein crowders are consistent with the conclusions from in-cell studies: weak, non-specific, attractive interactions play important roles in determining the effects of macromolecular crowding.

We close this section by noting that few significant changes in the chemical shift of backbone amide protons and nitrogens are observed in these studies, suggesting that even though the effect of non-specific interactions on stability can be considerable, they are individually transient and weak. This situation is unlike ligand binding, where specific binding to a particular domain of a protein can lead to large changes in chemical shift (56).

1.12 Enthalpy and entropy

In an effort to understand the contributions of hard- and soft- interactions, several groups have measured the temperature dependence of amide proton exchange to quantify the contribution of enthalpy and entropy to macromolecular crowding effects. Using a β-hairpin peptide as a test molecule, the Harries group (57, 58) observed stabilizing effects for small polyol osmolytes and the synthetic polymers, polyethylene glycol and dextran. The enthalpic components differed for the two classes of cosolute. Osmolyte-induced stabilization arose from an increased enthalpic contribution. Low concentrations of polymers also acted enthalpically, but higher polymer concentrations tended to drive the system entropically. Independent studies using larger test proteins with both synthetic polymers and proteins as crowders have also been reported. Both Wang et al. (59), who studied his-tagged ubiquitin, and Benton et al. (47) who used CI2, observed complicated relationships between the effects of crowding on $\Delta G_{den}^{o'}$ and $\Delta H_{den}^{o'}$ using ubiquitin as the test protein. That is, an increased

$\Delta G_{den}^{o'}$ was not always associated with a decreased $\Delta S_{den}^{o'}$ (and *vice versa*), plus the change in $\Delta H_{den}^{o'}$ was non-zero. As pointed out by Benton, Smith, Young and Pielak (47) some of the complication observed for synthetic polymers may arise from the influence of preferential hydration (8). In summary, the simplest ideas about macromolecular crowding predict entropically derived stabilization, but recent studies indicate that attractive, non-specific interactions complicate this interpretation.

1.13 Protein stability in cells

Interpretation of in-cell stability studies in terms of hard-core repulsions alone is problematic. Ghaemmaghami and Oas (60) studied the stability of λ repressor protein (8.8 kDa) in *E. coli* by using amide proton hydrogen exchange and mass spectroscopy. Their main conclusion was that stability is unchanged in cells. The Gierasch lab fluorescently tagged cellular retinoic acid-binding protein I (16 kDA) in cells and measured stability by using urea denaturation (61, 62). They showed that the test protein was destabilized in cells. In summary, these seminal studies prove that crowding in cells involves more than just hard-core interactions.

Next, we turn to studies that did not use perturbing cosolutes. The inability of the cytoplasm to fold ProtL was discussed above (38). Inomata, *et al.* (63) used in-cell NMR to assess the amide proton exchange rates for ubiquitin in HeLa cells. The rates increased, suggesting destabilization of the protein. The Gruebele group studied the stability of the test protein phosphoglycerate kinase in human cell lines by using fluorescence spectroscopy (64-66). Their main conclusion was that the folding kinetics, stability and dynamics are not only affected by crowding, but also by the environment in different cellular compartments.

Specifically, the protein folded faster and was stabilized in the nucleus compared to its behavior in the endoplasmic reticulum and the cytoplasm.

Modulating the strength of interactions is more straightforward *in silico* than it is in the laboratory, and combining the results from simulation with those from 'wet' experiments can yield important insight. A molecular model of the bacterial cytoplasm has been developed by McGuffee and Elcock (41). It depicts an atomistic view of the macromolecules inside *E. coli*, and the authors implemented Brownian dynamics simulations to assess the protein stability. They utilized the model to simulate the systems described above (60, 61) using four scenarios. The first scenario considered only hard-core repulsions and, as predicted, the result was stabilization. They then turned to a 'full energy' model, which incorporated electrostatic interactions and a Lennard-Jones potential to define steric, dispersion forces and hydrophobic interactions. The second scenario incorporated this full energy model for the cytoplasm, but used only hard-core repulsions for macromolecule-test protein interactions. A smaller stabilization effect was observed. The third scenario used hard-core repulsions between cytoplasmic proteins and the full model for macromolecule-test protein interactions. The result was destabilization. Lastly, they used the full energy model for both the cytoplasmic macromolecules and the test proteins, and found that their results match the experimental data.

In summary, these detailed simulations showed the importance of incorporating both and hard- and soft- interactions to obtain physiologically relevant information. Further experimental quantification of enthalpic and entropic contributions in crowded systems is expected to aid in the development of detailed and biologically relevant models that will be even more useful for simulating the biophysical effects of the crowded cellular environment.

1.14 Conclusions

Soft interactions are difficult to assess, yet it is essential that we understand them because they can diminish or even reverse the effects of hard-core repulsions. There are several reasons for the difficulty. First, soft interactions can be attractive or repulsive. Second, they arise from multiple phenomena: charge-charge interactions, hydrogen bonding, the hydrophobic effect, *etc*. Third, although individually weak, their final effect depends on the sum of individual components. Fourth, their weak and additive (and probably synergistic), nature presents a major challenge to developing the force fields required to drive simulations of crowding effects. Fifth, resolving and identifying these individual components from wet laboratory data is expensive. These are the challenges *in vitro*, where reductionist approaches can be applied. It is even more difficult to identify soft interactions in the heterogeneous cellular environment where reductionist approaches are difficult or impossible to apply. Nevertheless, endeavors aimed at quantifying soft interactions are essential for producing a physiologically relevant understanding of biophysics. Once the details of soft interactions are known, it should be possible to tune them so as to obtain bespoke behavior in test tubes and in cells. In summary, biologists of all types must keep these interactions in mind when designing experiments to correlate *in vitro* studies with *in vivo* behavior.

1.15 Figure

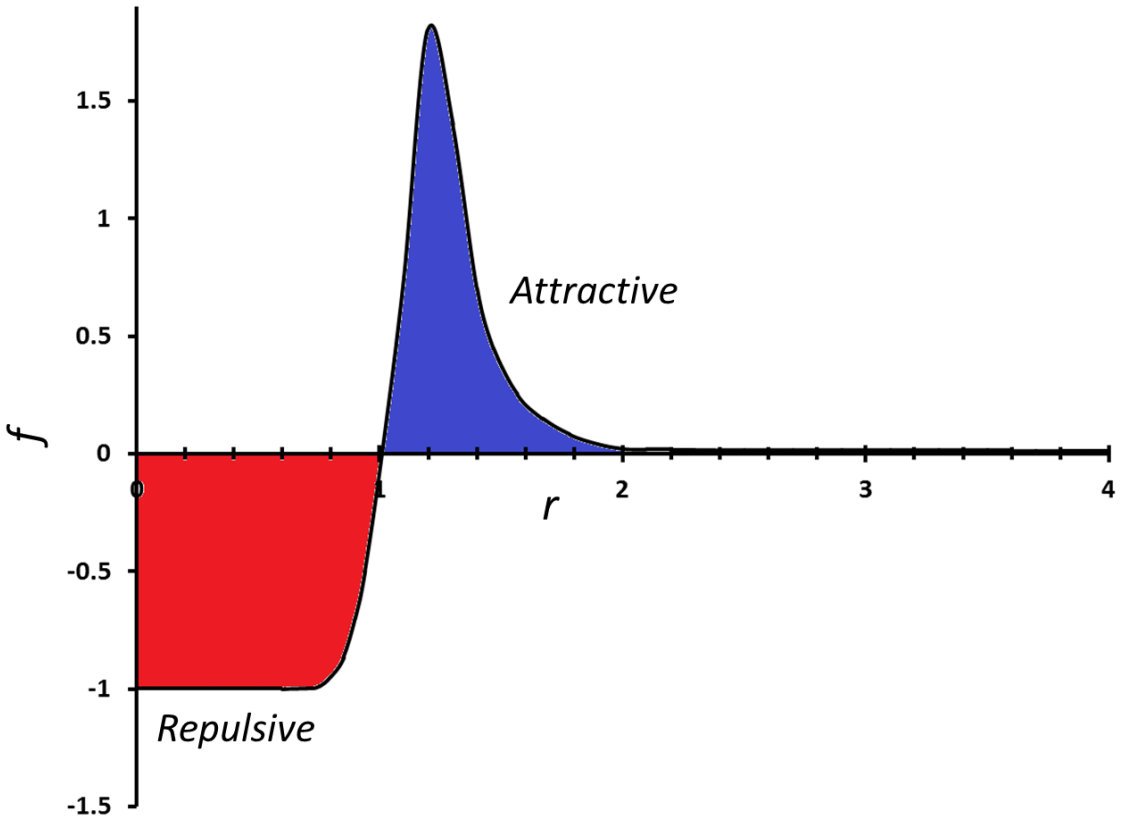

Figure 1.1: Normalized Mayer f-function (*f*) as a function of the distance (r, in arbitrary units) between two spheres. The sum of the shaded regions defines the excluded volume. The excluded volume can be made positive or negative by adjusting the strength of repulsive (red) and attractive (blue) interactions. Adapted from (14).

CHAPTER 2: IMPACT OF RECONSTITUTED CYTOSOL ON PROTEIN STABILITY

The material in this chapter is adapted from:

Sarkar M, Smith AE, Pielak GJ. Impact of reconstituted cytosol on protein stability. *Proceedings of National Academy of Sciences;* in press.

(M.S. and G.J.P. designed research; M.S. and A.E.S. performed research; M.S., A.E.S and G.J.P. analyzed data; M.S., A.E.S and G.J.P. wrote the paper.)

2.1 Introduction

Macromolecules in *Escherichia coli* reach concentrations of 300-400 g/L and occupy up to 40% of the cellular volume (1), but proteins are normally studied in buffer alone. The effects of crowding arise from two phenomena, hard-core repulsions and non-specific chemical (soft) interactions (16, 53, 59, 67-71). Hard-core repulsions limit the volume available to biological macromolecules for the simple reason that two molecules cannot be in the same place at the same time. This press for space favors compact states over expanded states. The second phenomenon arises because crowders not only exclude volume, but also participate in chemical interactions. Even though individually weak, the high concentration of macromolecules can lead to a large net effect. Repulsive non-specific interactions reinforce the hard-core repulsions, whereas attractive non-specific interactions oppose them. We use the term "non-specific attractive interactions" to distinguish these from specific chemical interactions, such as ligand binding.

Our aim is to understand how the crowded and heterogeneous, intracellular environment affects the equilibrium thermodynamic stability of globular proteins. Globular

proteins are marginally stable in buffer at room temperature (2), with Gibb's free energy differences of 5 to 15 kcal/mol between the efficiently-packed native (N) state and the ensemble of higher energy denatured (D) states ($\Delta G_D^{o\prime}$) (5). Crowding effects arise from entropic and enthalpic contributions; hard-core repulsions are entropic while the consequent non-specific chemical interactions are also enthalpic. Hard-core repulsions always increase $\Delta G_D^{o\prime}$ for globular proteins because D occupies more space than N (6, 15, 72). On the other hand, non-specific interactions can favor N or D, depending on their nature (67). Non-specific repulsions increase $\Delta G_D^{o\prime}$ because they make the crowder appear even larger. Non-specific attractions decrease $\Delta G_D^{o\prime}$ because more favorable interactions are able to form as the protein unfolds (16, 53, 59, 67-71).

Most efforts to understand macromolecular crowding use uncharged synthetic polymers as crowding agents (15, 67, 73). Such polymers tend to emphasize hard-core repulsions, and therefore stabilize globular proteins, but, they are not accurate physiological mimics because most synthetic polymers are non-globular and relatively inert (35, 73). Protein crowders are better mimics because cells contain large amounts of protein. Protein crowders can be destabilizing because they interact with proteins *via* attractive, weak non-specific contacts with the backbone (35, 53, 54). Nevertheless, cells are not crowded with one particular protein. For instance, an *E. coli* cell possesses 2.4 million protein molecules representing approximately 4000 different proteins (74, 75), resulting in a diverse population of charges, sizes and amino acid compositions.

The intracellular environment is optimal for understanding the physicochemical effects of the cytoplasm, but measuring stability in living cells is challenging because application of reductionist-based methods involving alteration of the cytosol are inconsistent

with viability (76). Studies of in-cell stability by urea denaturation suggest that the cytoplasm does not affect stability (60, 61), whereas results from McGuffee and Elcock in their pioneering molecular dynamics simulation of the *E. coli* cytoplasm (41) point to destabilization.

A non-perturbing method is clearly advantageous for quantifying the effect of the cellular interior. NMR-detected backbone amide proton exchange (77) is one such method and, unlike others, reports stability at the residue level. As discussed below, under certain specific conditions (78) the technique yields the free energy required to expose amide protons to solvent, $\Delta G^{o'}_{op}$. For a particular protein, the largest values of this parameter occur upon global unfolding and are equivalent to the denaturation free energy, $\Delta G^{o'}_{D}$ (78).

Detecting resonances from the backbone of globular proteins by NMR in *E. coli* is difficult because non-specific interactions limit rotational motion, broadening ^{15}N-^{1}H crosspeaks from heteronuclear single quantum coherence (HSQC) spectra into the background (28, 29, 32). Here, we use lyophilized *E. coli* cytosol (35, 79-81) to mimic the intracellular environment and quantify its effects on the stability of the test protein chymotrypsin inhibitor 2 (CI2), which is amenable to NMR-detected amide proton exchange (82). The cell extracts not only allow us to more directly assess the results of the molecular dynamics simulation (41), but also allows the application of a reductionist approach not possible using live cells in that the cytosol concentration can be varied.

We extracted the cytoplasm from saturated cultures and removed the low molecular weight components to observe only the macromolecular effects. After exchanging the labile protons with deuterons, we lyophilized the sample to obtain powdered cytosol. We used NMR-detected amide proton exchange (78) to obtain residue level stabilities of ^{15}N enriched

CI2 in solutions crowded with 100 g dry weight /L and 130 g dry weight /L reconstituted cytosol.

2.2 Results

2.2.1 Suitability of Reconstituted Cytosol

McGuffee and Elcock (41) used an inventory of *E. coli* proteins compiled by Link et al. (83) to design a model cytosol. We made lyophilized cytosol and used mass spectrometry and proteomics to identify the proteins. Our list of 233 proteins (Table 2.1) contains 33 of the 45 proteins from the model. The difference probably reflects the fact that McGuffee and Elcock used the list from growth on minimal media and we used cytosol from cells grown in rich media. We also assessed the ratio of the protein mass in the lyophilized cytosol to its dry weight. Triplicate Lowry analyses of the 100.0 and 130.0 g dry weight /L reconstituted samples showed they contained 52 ±4 g/L and 87 ±3 g/L protein, respectively. Thus, half to two-thirds of the mass of our lyophilized cytosol comprises proteins, the remainder being mostly nucleic acids, which is reasonable (74). These observations indicate that the results from its use are appropriate for comparisons to the results from McGuffee and Elcock's study.

2.2.2 Weak Interactions

For stable, globular proteins like CI2, NMR chemical shifts reflect the highly populated N state. Significant chemical shifts changes are expected if a protein interacts strongly with surrounding molecules (56). We studied the weighted chemical-shift difference from ^{15}N-^{1}H HSQC spectra (43, 47, 53, 84) of the I29A;I37H variant in 130.0 g dry weight /L sample and in buffer alone. Out of the 45 observable residues, only

five showed significant shift changes (Gly 10, Glu 26, Leu 32, Met 40 and Asn 56; Figure 2.1). Apart from Leu 32, these residues are in unstructured regions. A plausible explanation is that the cytosol induces small structural changes in flexible regions. The fact that only small chemical shift changes are observed throughout the rest of the protein indicates that CI2-crowder interactions are non-specific, weak and transient, in agreement with previous studies (35). This result also suggests that the N state of CI2 is largely unperturbed in reconstituted cytosol.

2.2.3 Measuring $\Delta G_{op}^{o'}$

Amide-proton exchange experiments provide equilibrium thermodynamic data about protein stability if four assumptions are valid. First, the test protein must be stable ($\Delta G_{op}^{o'} > 0$). Second, it must remain intact during the experiment. Third, the intrinsic rate of exchange, k_{int}, must be rate determining. Fourth, k_{int} values, which are calculated in buffer (85), must be applicable in the presence of co-solutes. These assumptions are true in buffer (86), in several synthetic crowders (43, 47) and in protein solutions (53). Here, we assessed these assumptions for reconstituted cytosol.

The cytosol does not induce unfolding or misfolding, because changes in CI2 backbone chemical shift are small (Figure 2.1). Nevertheless, the cytosol could cause CI2 to aggregate, and cytosolic proteases could degrade it. The signal decay from aggregation or proteolysis would compete with the signal decay from exchange, complicating interpretation. To test for these problems we performed a pseudo amide-proton exchange experiment using non-exchanged cytosol in H_2O at our highest concentration of reconstituted cytosol (130.0 g dry weight /L). There was no significant change in CI2 crosspeak intensities over 24 h (Table 2.2). These observations indicate that the first two assumptions are valid; CI2 remained

folded, intact and un-aggregated in the reconstituted cytosol for the duration of the amide ^{1}H exchange experiments.

The third assumption is that k_{int} is rate determining. Unfortunately, the methods commonly used to verify this assumption (28) -- changing the pH and NOESY-detected amide proton exchange (78, 86, 87) -- do not work in cytosol. The pH change method fails because CI2 stability depends on pH (53) and the reconstituted cytosol aggregates upon altering the pH. Nor did we observe crosspeaks in a 30 min, time-resolved NOESY-detected amide proton exchange spectrum. Increasing the time gave more signal, but the diagonal crosspeaks were lost due to the exchange. Instead, we turned to a thermodynamic cycle (88) to ensure we are measuring $\Delta G_{op}^{o'}$ (Figure 2.2).

The cycle comprises two CI2 variants, I29A;I37H and I57A;I37H, and two solvent conditions, dilute and reconstituted cytosol at 100.0 g dry weight /L. These variants were chosen because they allow us to observe complete amide proton exchange within 24 h. We calculated $\Delta G_{op}^{o\prime}$ assuming k_{int} is rate determining. If the assumption is correct, the effect of crowding on the variants (horizontal sides) should be equal, as should the effect of the mutation on the solvent conditions (vertical sides). These conditions are met (Figures 2.3 and 2.4), suggesting that we are measuring $\Delta G_{op}^{o'}$. This result could be fortuitous, but two observations support its veracity. First, we know that k_{int} is rate determining for the I29A;I37H variant in buffer alone, in 100.0 g/L polyvinylpyrrolidone (PVP), bovine serum albumin (BSA) and lysozyme at pH 6.5, 20 ^{o}C (42, 43, 53). Second, the folding rates of the I29A and I57A variants are 11 s^{-1} and 31 s^{-1} (89), much larger than the intrinsic rates (~1 s^{-1}).

The final assumption is that the cytosol does not change k_{int}. Even though the disordered loop (residues 35-45) is not observed in our exchange experiments, due to rapid

solvent exchange, it serves as an internal control for assessing the effect of the cytosol. Previously, we used the loop to show that a synthetic polymer and protein crowders do not change k_{int} (42, 43, 53). This assumption was tested for the cytosol by applying the $^{15}N^{H/D}$ SOLEXSY experiment (90). We found that the intrinsic exchange rates remain unchanged in reconstituted cytosol (79).

2.3 Discussion

NMR detected amide proton exchange (77, 78) allows quantification of protein stability at the level of individual residues. We measured the stability, compared to buffer alone, at two cytosol concentrations at pH 6.5 and 20 ^{o}C. The $\Delta\Delta G_{op}^{o'}$ values (cytosol minus buffer) for the I29A;I37H variant are superimposed on the backbone of CI2 in Figure 2.5. Increasing the cytosol concentration from 100.0 g dry weight /L to 130.0 g dry weight /L, decreases the stability further, especially in α-helix and β-sheet regions. Even though reconstituted cytosol is destabilizing, a histograms of $\Delta G_{op}^{o'}$ *versus* residue number (Figure 2.6) shows similar trends in reconstituted cytosol and buffer alone, suggesting that the folding mechanism is unchanged. We were unable to quantify $\Delta G_{op}^{o'}$ in certain residues for two reasons. First, some are in unstructured regions (*e.g.*, the loop, residues 35 - 45). These amide protons exchange too fast to assess rates. Second, the cytosol broadens some resonances beyond detection.

It is tempting to correlate structural features of CI2 with the observed destabilization, but interpreting amide proton exchange data in such detail remains controversial. For instance, it has been argued that caution should be used in attempts to correlate rates with solvent accessible area (91-94). In addition, CI2 possesses fast exchanging amide protons in

buried regions and slow exchanging protons in non-hydrogen bonded regions (82). In particular, its α-helix contains both partially (Ala 16 and Lys 24) and completely buried (Ile 20 and Leu 21) residues (82). These regions are marginally stabilized in 100.0 g dry weight/L reconstituted cytosol (<0.3 kcal/mol) but destabilized at the higher concentration. Maximum destabilization was observed in the β strand comprising Asp 45 to Asp 52, and most of these residues are buried in that part of the core that exchanges only upon global unfolding (82). Among the residues we observe, only Val 34, is significantly stabilized (> 0.3 kcal/mol) at both concentrations. Interestingly, this residue is immediately before the extended loop whose structure, as we described above, may be affected in cytosol. We did not observe a change in $\Delta G_{op}^{o'}$ for Val 34 in our previous studies using PVP, BSA or lysozyme (53) as crowding agents. One explanation could be that components of the reconstituted cytosol interact specifically with this region. We looked for correlations between changes in stability and solvent accessible surface area and hydrogen bond lengths, but have found none (Figures 2.7 and 2.8). Taken together, these observations support both the idea that the interactions between components of the reconstituted cytosol and the protein backbone are mostly non-specific, and as discussed next, we focus on positions that exchange only by global unfolding.

We averaged the $\Delta\Delta G_{op}^{o'}$ for CI2 residues that exchange on global unfolding (82), and compared (Figure 2.5) the effects of the reconstituted cytosol to the effects of inert synthetic polymers and biologically relevant protein crowders (47, 53). Ficoll shows maximum stabilization followed by PVP. The more biologically relevant crowders BSA, lysozyme and cytosol destabilize CI2. We also know that Ficoll and PVP interact only weakly with the protein, whereas the biologically relevant crowders have stronger interactions (34, 35). We

conclude that transient, non-specific interactions between the test protein and other proteins cause destabilization.

Although our reconstituted cytosol is a reasonable model for testing McGuffee and Elcock's simulation (41) it is not a perfect mimic of the intracellular environment. First, the DNA is sheared during preparation and the mRNA has been degraded. Second, *E. coli* contains a large number of small molecules (95-97), but our reconstituted cytosol was dialyzed against water, leaving only the ions required for charge neutralization. Many of the absent small molecules, especially osmolytes (98), affect protein stability, and we have shown that salt mitigates the destabilizing effect of crowding by BSA (53). Thus, a unified theory of intracellular crowding cannot be deduced from our reconstituted cytosol. Perhaps the absence of small molecules explains the difference between our observation of destabilization and findings from in-cell studies where no change in stability was observed (60, 61). We note, however, that small changes in the conditions used in those studies lead to the conclusion that the cytoplasm is slightly destabilizing .

Despite these shortcomings, reconstituted cytosol is probably a better mimic than homogenous crowders such as Ficoll, PVP, BSA, *etc*., and a similar preparation has been successfully used to study changes in side chain dynamics of calmodulin (80, 81). Reconstituted cytosol provides an intermediate step between dilute solution and in-cell experiments. Moreover, direct residue-level NMR studies of globular proteins are not possible in *E. coli* because, with a few exceptions (32), it is difficult to observe high quality HSQC spectra from proteins in *E. coli* cells (28, 29, 32).

Our residue-level analysis validates the cytoplasmic model (41), and our results are generally consistent with investigations of global protein stability in cells (60, 61). All these

studies indicate that intracellular environment offers enough non-specific interactions to overcome the stabilizing effects of hard-core repulsion. Our results highlight the need to incorporate non-specific, weak interactions in the modeling of macromolecular crowding.

Traditionally, investigations of macromolecular crowding effects focused on hard-core repulsions, which predict only stabilization. Our result contradicts these predictions. The reason may be that most studies used synthetic polymers, such as Ficoll and Dextran. These polymers have only weak interactions, whereas proteins and cytosol have stronger non-specific interactions with CI2 (34, 35). Synthetic, homogenous crowding systems have provided fundamental insight into macromolecular crowding, but, as stated by Elcock (73), they may not resemble the intracellular environment because they cannot mimic the interactions in the cytosol. Only recently has the importance of both non-specific, attractive interactions and steric repulsions been recognized (16, 53, 59, 67-71). Furthermore, a new idea about the temperature dependence of crowding effects suggests that there is a crossover temperature where the stabilizing and destabilizing effects cancel (99). Clearly, combining experimental and modeling efforts will increase our understanding of how the intracellular environment affects protein stability.

Non-specific interactions arise from several sources. For example, we know that coulombic interactions (53), hydrogen bonds and the hydrophobic effect are key players (32). These interactions are probably of functional importance in cells. For instance, it has been hypothesized that weak protein-protein interactions give rise to cellular organization (100), as observed by centrifugation of whole cells (101) and in protein clusters (102). Such clustering leads to the formation of 'supercrowded' regions surrounded by zones of dilute cytoplasm, which help direct cellular metabolism (103). In summary, our results provide strong evidence

for the importance of non-specific attractive interactions in determining protein stability in cells, and highlight the significance of using reconstituted cytosol to mimic cellular crowding.

2.4 Materials and Methods

2.4.1 Test proteins

The CI2 gene was mutated and the variants expressed and purified as described (42, 53). ^{1}H and ^{15}N chemical shift assignments for the ^{15}N and ^{13}C enriched I57A;I37H variant were obtained in dilute solution (50 mM sodium phosphate, pH 6.5, 20 ^{o}C) by using the HNCACB (104) and CBCA(CO)NH (105) experiments on a 600 MHz Varian Inova spectrometer equipped with a triple resonance probe. The ^{1}H, ^{13}C and ^{15}N spectral widths were 11990 Hz, 12064 Hz and 2500 Hz, respectively. The HNCACB spectrum was acquired with 60 and 32 complex increments and the CBCA(CO)NH spectrum was acquired with 48 and 24 complex increments in the ^{13}C and ^{15}N dimensions, respectively. The number of complex points in the ^{1}H dimension was 1024. Data were processed with NMRpipe (106) and NMRView (107). Assignments for the I57A;I37H variant are given in Table 2.6. Assignments for I29A;I37H variant are from (84).

2.4.2 Reconstituted cytosol

The preparation of reconstituted cytosol has been described (35, 79), but in this instance a 28 L culture was used. The lyophilized and deuterated cytosol (1.20 g) was suspended in enough deuterated 50 mM sodium phosphate buffer to give a 12.0 mL of solution. The pH_{read} [i.e., uncorrected for the isotope effect (108)] was adjusted to 6.5. The solution was centrifuged at 14000 *g* for 40 min at 4 ^{o}C. The supernatant, which contained

100.0 g dry weight /L of reconstituted cytosol was used for six trials: three trials each on the I29A;I37H and I57A;I37H variants. A small portion of the reconstituted cytosol was retained for analysis (*vide infra*).

2.4.3 NMR detected amide proton exchange of crowded solutions by reconstituted cytosol

Amide proton exchange experiments were performed as described (28, 53). Prior to starting an experiment, the shims were optimized with a 1 mL, 1 mM CI2 in 50 mM sodium phosphate, pH 6.5, containing 10% (v/v) D_2O (78). The spectra were acquired with 1024 complex points in the ^{1}H dimension and 48 complex increments in the ^{15}N dimension. Sufficient ^{15}N enriched CI2 was added to 1 mL of the deuterated, reconstituted cytosol to give a final concentration of 1 mM CI2. The solution was immediately transferred to a 5 mm NMR tube (Norell) and 22 serial HSQC datasets were acquired at 20 ^{o}C using the 600 MHz spectrometer. Exchange experiments were performed thrice to obtain the mean and its standard deviation. The intrinsic rates of amide exchange (k_{int}, s^{-1}) were obtained from the online platform SPHERE (85) at pH_{read} 6.5, 20 ^{o}C in 50 mM sodium phosphate buffer and 100% D_2O, because the cytosol does not change intrinsic rates (79).

The I29A;I37H variant was also studied at a higher cytosol concentration, 130 g dry weight /L. These experiments were performed on a 500 MHz Varian Inova spectrometer with a triple resonance HCN cold probe with ^{1}H and ^{15}N spectral widths of 8401.6 Hz and 2200 Hz, respectively.

2.4.4 Characterization of reconstituted cytosol

A modified Lowry assay (Thermo Scientific) was performed after 100-fold dilution of the two reconstituted cytosol samples with NMR buffer (50 mM sodium phosphate, pH

6.5). The total protein concentration was obtained from the absorbance at 750 nm by using a standard curve made from 0.25, 0.50, 1.0 and 1.5 mg/ml solutions of bovine serum albumin. The assay was performed thrice to obtain a mean and its standard deviation. The samples contained 52 ±4 g/L and 87 ±3 g/L proteins, respectively.

Mass spectrometric analysis was performed at the UNC Michael Hooker Proteomics Center. The lyophilized cytosol was resuspended in a solution containing 8 M urea and 100 mM NH_4HCO_3 and reduced for 30 min with 10 mM dithiothreitol. The proteins were then acetylated with 40 mM iodoacetamide in the dark for 30 min. The reaction was quenched by adjusting the reducant concentration to 40 mM. The urea concentration was reduced to 0.5 M by dilution with 100 mM NH_4HCO_3, and the sample was digested with trypsin (1:10 enzyme:protein by weight) overnight at 37 °C. The peptides were purified by using C-18 Spin Columns (Pierce Biotechnology), and lyophilized. The peptides were suspended in 5% acetic acid and loaded onto a 2 cm × 360 µm o.d. × 100 µm i.d. microcapillary fused silica precolumn packed with Magic 5 µm C18AQ resin (Michrom Biosciences). The loaded precolumn was washed with 95% Solvent A (0.1% aqueous formic acid) /5% Solvent B (0.1% formic acid in acetonitrile) for 20 min at a flow rate of 2 µL/min and then connected to a 360 µm o.d. × 75 µm i.d. analytical column packed with 14 cm of 5 µm C18 resin constructed with an integrated electrospray emitter tip. The peptides were eluted at a flow rate of 250 nL/min by increasing solvent B to 40% with a Nano-Acquity HPLC solvent delivery system (Waters) The system was directly connected through an electrospray ionization source interfaced to a LTQ Orbitrap XL ion trap mass spectrometer (Thermo Fisher Scientific). The mass spectrometer was controlled with Xcalibur software and operated in the data-dependent mode, in which the initial scan recorded the mass-to-charge

ratios of ions from 400 to 2000. The ten most abundant ions were automatically selected for subsequent collision-activated dissociation. Raw files were searched by using MASCOT software (Matrix Science, Ver. 2.3.02) *via* Proteome Discoverer (Thermo., Ver. 1.3.0.339) against the UniProt *E. coli* reference database (109). Search parameters included a peptide mass tolerance of 10 ppm, fragment ion tolerance of 0.8 Da, and variable modifications for methionine oxidation and carbamidomethylation of cysteine. Identification of two or more peptides for a given protein was considered sufficient evidence to state the protein was present. Proteins with more than two identified peptides are listed in Table 2.1

2.5 Tables and figures

Table 2.1: Analysis of reconstituted cytosol. Proteins with two or more identified peptides along with the percent sequence coverage are listed. Ranks are from Link et al. (83) for growth in rich media. Unranked proteins identified in both analyses are marked an X. Proteins used by McGuffee and Elcock (41) are highlighted in yellow. We identified 233 individual proteins, of which 33 match the 45 relevant proteins used by McGuffee and Elcock.

Peptides	%Coverage	Names	Gene	Rank
25	42.25	Tryptophanase		4
25	41.62	elongation factor Tu	TufA	X
23	47.02	molecular chaperone DnaK	DnaK	X
21	34.66	elongation factor G	FusA	117
19	37.23	molecular chaperone GroEL	Mop	52
17	36.80	30S ribosomal protein S1		86
16	33.89	phosphoenolpyruvate carboxykinase	PckA	121
14	42.62	fructose-bisphosphate aldolase class 2	Fba	93
14	13.56	DNA-directed RNA polymerase subunit beta	Rpo	
12	38.76	phosphoglycerate kinase	Pgk	31
12	35.29	cysteine synthase A	CysK	X
11	30.91	citrate synthase		
11	28.61	isocitrate dehydrogenase	IcdA	X
11	13.22	DNA-directed RNA polymerase subunit beta'	Rpo	X
10	48.08	malate dehydrogenase	Mdh	21
10	30.51	glyceraldehyde-3-phosphate dehydrogenase A	GapA	X
10	30.32	trigger factor	Tig	44
10	20.68	dihydrolipoyl dehydrogenase	LpdA	25
10	16.93	glycine dehydrogenase		
10	16.68	carbamoyl-phosphate synthase large chain		
9	51.87	alkyl hydroperoxide reductase subunit C	AhpC	28
9	46.72	DNA-binding protein H-NS	Hns	
9	18.32	molecular chaperone ClpB		
9	15.78	pyruvate dehydrogenase E1 component		55
8	42.38	glutamate/aspartate periplasmic-binding protein		
8	31.20	2,3-bisphosphoglycerate-dependent phosphoglycerate mutase	GpmA	104
8	20.03	chaperone protein HtpG		X
8	16.25	catalase-peroxidase		

8	8.64	aconitate hydratase 1		
7	33.18	isocitrate lyase		
7	31.84	osmotically-inducible protein Y		39
7	28.62	molecular chaperone Hsp31 and glyoxalase 3		
7	27.08	enolase	Eno	77
7	24.48	succinyl-CoA ligase [ADP-forming] subunit beta	SucC	X
7	23.84	transcription elongation protein NusA		X
7	20.07	periplasmic oligopeptide-binding protein		22
7	16.10	acetyl-coenzyme A synthetase		
7	13.61	2-oxoglutarate dehydrogenase E1 component	SucA	X
6	57.78	DNA-binding protein HU-alpha	Hup	X
6	55.24	peroxiredoxin OsmC		16
6	31.55	transaldolase B		73
6	26.86	elongation factor Ts	Tsf	36
6	21.17	fumarate hydratase class I, aerobic		
6	20.54	succinate-semialdehyde dehydrogenase		
6	19.30	F0F1 ATP synthase subunit alpha		24
6	17.57	phosphoenolpyruvate-protein phosphotransferase	PtsI	X
6	15.66	aspartate aminotransferase		101
6	15.08	dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex	AceF	35
6	14.47	formate acetyltransferase 1		
5	33.33	flavoprotein WrbA		
5	33.08	histidine-binding periplasmic protein		59
5	28.48	universal stress protein E		
5	21.64	UDP-4-amino-4-deoxy-L-arabinose--oxoglutarate aminotransferase		
5	20.64	pyruvate kinase I		
5	20.51	6-phosphogluconate dehydrogenase, decarboxylating		
5	20.37	adenylosuccinate synthetase	PurA	X
5	20.17	phosphoserine aminotransferase	SerC	X
5	20.15	phosphopentomutase		
5	20.04	gamma-aminobutyraldehyde dehydrogenase		
5	17.95	50S ribosomal protein L1		3
5	14.89	Enoyl-[acyl-carrier-protein] reductase	FabI	X
5	14.15	serine hydroxymethyltransferase	GlyA	X
5	13.11	proline--tRNA ligase		
5	10.11	aminopeptidase N		
5	9.65	aldehyde-alcohol dehydrogenase		X
5	8.84	phosphoenolpyruvate synthase		
5	8.45	alanine--tRNA ligase		
4	70.16	acidic protein msyB		
4	33.54	bacterioferritin		

4	28.81	30S ribosomal protein S13		
4	25.87	nucleoside diphosphate kinase		
4	25.62	50S ribosomal protein L7/L12		49
4	24.83	50S ribosomal protein L9		34
4	24.06	cystine-binding periplasmic protein	FliY	89
4	23.58	stringent starvation protein A		X
4	23.30	30S ribosomal protein S10		32
4	21.48	FKBP-type peptidyl-prolyl cis-trans isomerase FkpA		67
4	21.09	uncharacterized oxidoreductase YghA		
4	20.00	triosephosphate isomerase	TpiA	98
4	18.24	DNA-directed RNA polymerase subunit alpha	Rpo	48
4	17.67	argininosuccinate synthase		X
4	17.45	2,5-diketo-D-gluconic acid reductase A		
4	17.23	aerobic respiration control protein	ArcA	X
4	15.89	S-adenosylmethionine synthase		
4	15.81	serine--tRNA ligase		
4	15.43	fructose-bisphosphate aldolase class 1	Fba	93
4	14.79	pyruvate kinase II		X
4	13.99	lactaldehyde dehydrogenase		
4	13.84	ABC transporter periplasmic-binding protein YtfQ		
4	10.49	methionine--tRNA ligase		
4	10.29	oligopeptidase A		
4	8.95	leucine--tRNA ligase		
4	8.70	transketolase 2		
3	62.35	phosphocarrier protein HPr	PtsH	7
3	53.85	acyl carrier protein		
3	33.33	DNA-binding protein HU-beta	Hup	X
3	32.14	thiol peroxidase		74
3	31.82	50S ribosomal protein L22		
3	26.63	glucose-specific phosphotransferase enzyme IIA component		51
3	26.56	phosphoheptose isomerase		
3	25.96	uracil phosphoribosyltransferase	Upp	X
3	25.31	Ecotin		
3	25.24	6-phosphofructokinase isozyme 2		
3	25.00	50S ribosomal protein L15		
3	24.64	cold shock-like protein CspC	CspC	X
3	23.64	50S ribosomal protein L10		
3	22.70	ribosome-recycling factor	Frr	
3	21.89	30S ribosomal protein S3		
3	21.89	50S ribosomal protein L4		41
3	20.90	50S ribosomal protein L6		10
3	20.11	30S ribosomal protein S7		
3	20.00	translation initiation factor IF-3		

3	19.62	transaldolase A		
3	18.99	50S ribosomal protein L5		
3	17.85	protein RecA		
3	17.65	succinyl-CoA ligase subunit alpha	SucD	X
3	16.23	NADPH-dependent curcumin reductase		
3	15.71	mannitol-1-phosphate 5-dehydrogenase		
3	15.56	glutathione reductase		
3	15.55	pantothenate synthetase		
3	13.73	soluble pyridine nucleotide transhydrogenase		
3	13.50	acetate kinase		
3	13.43	glutamine synthetase	GlnA	X
3	12.84	dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex	SucB	X
3	12.30	carbamoyl-phosphate synthase small chain		X
3	11.51	ATP-dependent protease ATPase subunit HslU		
3	10.87	ATP synthase subunit beta		12
3	10.30	lysine--tRNA ligase		
3	9.88	NADP-dependent malic enzyme		
3	9.25	galactitol-1-phosphate 5-dehydrogenase		
3	9.21	aspartate ammonia-lyase		
3	9.16	glycerol kinase		
3	9.01	asparagine--tRNA ligase		X
3	8.96	glucose-6-phosphate 1-dehydrogenase		
3	8.50	L-threonine 3-dehydrogenase		
3	8.39	pyruvate dehydrogenase		
3	7.55	glucosamine--fructose-6-phosphate aminotransferase		
3	6.98	dihydroxy-acid dehydratase		
3	6.20	valine--tRNA ligase		X
3	6.08	isoleucine--tRNA ligase		
3	4.83	translation initiation factor IF-2		
3	6.93	lysine--tRNA ligase		
2	48.19	glutaredoxin-3		
2	37.17	uncharacterized protein YebY		
2	32.98	galactitol-specific phosphotransferase enzyme IIB component		
2	30.36	uncharacterized protein YifE		X
2	29.21	uncharacterized protein YihD		
2	28.99	uncharacterized protein YjbJ		1
2	28.44	thioredoxin-1		
2	26.18	Fe/S biogenesis protein NfuA		
2	24.62	uncharacterized protein YdeI		
2	23.16	ribosome hibernation promoting factor		
2	22.92	universal stress protein A	UspA	30

2	22.06	50S ribosomal protein L16		
2	21.13	50S ribosomal protein L13		
2	20.73	leucine-responsive regulatory protein		
2	19.82	enhancing lycopene biosynthesis protein 2		
2	19.72	KHG/KDPG aldolase		
2	19.23	putative peroxiredoxin	Bcp	X
2	19.23	50S ribosomal protein L24		
2	18.50	superoxide dismutase [Cu-Zn]		
2	18.23	transcription antitermination protein NusG		
2	17.07	uncharacterized protein YfbU		
2	16.75	outer-membrane lipoprotein carrier protein		
2	16.10	50S ribosomal protein L20		
2	15.53	30S ribosomal protein S4		
2	15.38	uncharacterized protein YcaC		
2	14.81	putative ABC transporter arginine-binding protein 2	ArtI	71
2	14.66	uncharacterize protein YceI		110
2	14.37	uncharacterized protein YbeL		
2	14.34	3-oxoacyl-[acyl-carrier-protein] reductase FabG		
2	14.29	HIT-like protein hinT		
2	14.29	chaperone protein skp		
2	13.95	glutaredoxin-2		
2	13.85	30S ribosomal protein S9		
2	13.45	S-ribosylhomocysteine lyase		
2	13.23	glutaminase 1		
2	13.18	acetylglutamate kinase		
2	13.09	NH(3)-dependent NAD(+) synthetase	NadE	X
2	12.90	glutamine-binding periplasmic protein	GlnH	54
2	12.56	transcriptional regulatory protein PhoP		
2	12.50	glyoxylate/hydroxypyruvate reductase A		
2	12.40	imidazole glycerol phosphate synthase subunit HisF		
2	12.38	methionyl-tRNA formyltransferase		
2	12.20	3'(2'),5'-bisphosphate nucleotidase		
2	12.09	aminomethyltransferase	GcvV	103
2	11.81	phosphoribosylaminoimidazole-succinocarboxamide synthase		
2	11.38	3-oxoacyl-[acyl-carrier-protein] synthase 2		
2	11.33	branched-chain-amino-acid aminotransferase		
2	11.31	2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltransferase	DapD	X
2	11.31	quinone oxidoreductase		
2	11.30	33 kDa chaperonin		
2	11.19	uncharacterized protein YniA		
2	11.11	integration host factor subunit alpha		
2	11.11	ribose-phosphate pyrophosphokinase		
2	11.03	3-mercaptopyruvate sulfurtransferase	SseA	124

2	10.44	acetylornithine deacetylase		
2	10.20	alpha-galactosidase		
2	9.98	CDP-diacylglycerol--serine O-phosphatidyltransferase		
2	9.95	D-amino acid dehydrogenase small subunit		
2	9.86	50S ribosomal protein L11		
2	9.59	Gamma-glutamyl phosphate reductase		
2	9.57	50S ribosomal protein L3		
2	9.32	thymidine phosphorylase		
2	8.88	thiosulfate-binding protein	CysP	X
2	8.72	aspartate-semialdehyde dehydrogenase	Asd	
2	8.41	putative glyceraldehyde-3-phosphate dehydrogenase C		
2	8.38	GMP synthase [glutamine-hydrolyzing]		
2	7.96	betaine aldehyde dehydrogenase		
2	7.88	ferritin-1		
2	7.44	signal recognition particle receptor FtsY		
2	7.11	protein GrpE		
2	7.01	glutamate--tRNA ligase		
2	6.79	3-phosphoshikimate 1-carboxyvinyltransferase		
2	6.62	bifunctional purine biosynthesis protein PurH		X
2	6.01	biotin carboxylase	AccC	X
2	5.97	transcription termination factor Rho		X
2	5.84	phosphoglucosamine mutase		
2	5.81	glucose-1-phosphatase	Agp	54
2	5.77	uncharacterized ABC transporter ATP-binding uncharacterized protein YjjK		
2	5.61	adenylate kinase	Adk	92
2	5.47	aldehyde dehydrogenase B		
2	5.43	transketolase 1		
2	5.30	ketol-acid reductoisomerase	IlvC	X
2	5.08	N-succinylglutamate 5-semialdehyde dehydrogenase		
2	5.03	3-octaprenyl-4-hydroxybenzoate carboxy-lyase		
2	4.86	periplasmic dipeptide transport protein	DppA	102
2	4.64	polyribonucleotide nucleotidyltransferase		
2	4.24	aspartate--tRNA ligase		
2	4.20	DNA polymerase I		
2	3.98	DNA gyrase subunit B		
2	3.76	maltodextrin phosphorylase		
2	3.44	ribonuclease R		
2	3.40	phenylalanine--tRNA ligase beta subunit		
2	2.63	malate synthase G		

Proteins used by McGuffee and Elcock that we did not identify and their rank for growth in rich media. Unranked proteins are marked with an X.

elongation factor P (Efp)		X
glutamate synthase 4Fe-4S protein, small subunit (GltD)	X	
inorganic pyrophosphatase (Ppa)		68
peptidyl-prolyl cis-trans isomerase B (PpiB)		97
aspartate carbamoyltransferase (PyrI/PyrB)		X
ribose-5-phosphate isomerase (RpiA)	X	
superoxide dismutase, Mn (SodA)		123
5-methyltetrahydropteroyltriglutamate-homocysteine S-methyltransferase (MetE)	X	
superoxide dismutase, Fe (SodB)		40
dihydrodipicolinate synthase (DapA)		X
3-methyl-2-oxobutanoate hydroxymethyltransferase (PanB)	X	
polynucleotide phosphorylase (Pnp)		X

Table 2.2: Lack of decay shows that CI2 remains intact over the course of amide proton exchange experiments. A pseudo exchange was performed using non-deuterated reconstituted cytosol (130.0 g dry weight /L, pH 6.5, 20 ^{o}C). Twenty-two serial HSQC spectra were collected over a period of 24 h. SDev, standard deviation of the exponential fit from NMRView (107).

Residue	k_{obs} (s^{-1})	SDev	Residue	k_{obs} (s^{-1})	SDev
Trp5	8.10E-07	3.80E-07	Glu41	1.70E-07	4.70E-07
Val9	6.70E-07	3.70E-07	Tyr42	-5.90E-07	3.50E-07
Lys11	-2.50E-07	3.00E-07	Arg43	-3.90E-07	2.80E-07
Val13	2.10E-07	4.60E-07	Ile44	6.00E-07	9.00E-07
Ala16	1.60E-06	8.00E-07	Asp45	-8.00E-07	7.00E-07
Lys18	2.10E-07	2.90E-07	Arg46	4.70E-07	2.60E-07
Val19	-1.80E-07	3.40E-07	Val47	6.00E-07	5.00E-07
Leu21	7.80E-07	3.80E-07	Arg48	4.80E-07	3.60E-07
Gln22	8.00E-08	3.60E-07	Leu49	1.50E-07	3.20E-07
Asp23	1.03E-06	2.70E-07	Phe50	9.80E-07	3.00E-07
Lys24	-3.00E-07	3.00E-07	Val51	5.20E-07	3.40E-07
Ala27	-2.00E-07	6.00E-07	Asp52	-3.60E-06	3.70E-06
Val34	7.00E-07	5.00E-07	Asp55	6.00E-07	9.00E-07
Gly35	9.00E-06	1.60E-05	Asn56	6.10E-07	1.80E-07
Thr36	-1.10E-06	1.00E-06	Ala57	9.80E-07	3.00E-07
Ile37	1.30E-07	3.40E-07	Ala58	1.66E-06	4.10E-07
Val38	3.30E-07	3.40E-07	Glu59	5.70E-07	2.50E-07
Thr39	-1.80E-06	6.00E-07	Arg62	7.20E-07	2.40E-07
Met40	-2.00E-07	5.00E-07	Val63	1.01E-06	2.20E-07

Table 2.3: $\Delta G^{o'}_{op}$, in kcal/mol, for the I29A;I37H and I57A;I37H variants (pH 6.5, 20 ^{o}C) in 100.0 g dry weight /L of reconstituted cytosol. Protons that exchange upon global unfolding (82) are bolded. SEM; standard error of the mean.

	I29A;I37H				I57A;I37H			
Residue	**Trial 1**	**Trial 2**	**Trial 3**	**Average ±SEM or range**	**Trial 1**	**Trial 2**	**Trial 3**	**Average ±SEM or range**
Trp5	6.38	6.28	6.63	6.4 ±0.1	6.12	6.18	6.06	6.12 ±0.03
Glu7	_	6.28	_	6.28	5.23	5.87	_	5.6 ±0.6
Leu8	6.16	5.79	5.94	5.9 ±0.1	5.48	5.86	5.02	5.5 ±0.3
Val9	6.14	5.82	5.84	5.9 ±0.1	5.04	5.88	6.09	5.7 ±0.3
Gly10	6.20	6.11	_	6.2 ±0.1	_	_	_	_
Lys11	**7.14**	**6.99**	**7.00**	**7.04 ±0.05**	**6.88**	**6.27**	**6.67**	**6.6 ±0.2**
Val13	6.14	6.14	6.22	6.17 ±0.03	5.92	5.59	5.38	5.6 ±0.2
Glu15	_	_	_	_	4.08	3.98	3.75	3.9 ±0.1
Ala16	7.29	7.11	7.19	7.19 ±0.05	7.35	7.52	7.63	7.5 ±0.1
Lys17	5.41	5.39	5.42	5.41 ±0.01	6.69	6.87	6.93	6.8 ±0.1
Lys18	6.22	6.22	6.23	6.22 ±0.00	6.54	6.99	7.21	6.9 ±0.2
Val19	_	6.33	_	6.33	5.50	6.85	_	6.2 ±1.4
Leu21	**_**	**5.68**	**_**	**5.68**	**5.82**	**6.16**	**6.09**	**6.02 ±0.10**
Gln22	6.90	6.74	6.68	6.8 ±0.1	6.20	6.92	7.14	6.8 ±0.3
Asp23	_	6.68	_	6.68	_	_	_	_
Lys24	6.19	6.08	6.22	6.16 ±0.04	6.46	6.39	6.74	6.5 ±0.1
Ala27	6.02	5.97	5.95	5.98 ±0.02	6.29	6.33	6.36	6.33 ±0.02
Leu32	6.04	5.54	6.08	5.9 ±0.2	_	_	_	_
Val34	4.37	4.27	4.18	4.3 ±0.1	4.91	4.95	4.96	4.94 ±0.01
Arg46	6.80	6.67	6.56	6.7 ±0.1	6.25	6.49	6.35	6.4 ±0.1
Val47	**6.34**	**6.06**	**6.33**	**6.2 ±0.1**	**5.61**	**6.00**	**5.52**	**5.7 ±0.2**
Arg48	6.87	6.72	6.62	6.7 ±0.1	6.26	6.77	6.83	6.6 ±0.2
Leu49	**6.33**	**6.66**	**6.31**	**6.43 ±0.02**	**6.03**	**6.37**	**6.11**	**6.2 ±0.1**
Phe50	**6.51**	**6.34**	**6.17**	**6.3 ±0.1**	**6.10**	**5.88**	**6.07**	**6.0 ±0.1**
Val51	**_**	**6.87**	**6.82**	**6.85 ±0.05**	**_**	**_**	**_**	**_**
Asp52	6.44	6.40	6.45	6.43 ±0.02	5.90	_	6.58	6.2 ±0.7
Asn56	7.64	7.07	7.25	7.32 ±0.17	_	_	_	_
Ala57/Ile57	_	6.86	6.67	6.8 ±0.2	6.77	7.20	7.26	7.1 ±0.2
Ala58	7.50	6.48	6.51	6.8 ±0.3	6.31	6.80	7.27	6.8 ±0.3
Glu59	_	7.27	7.25	7.26 ±0.01	6.48	7.07	6.80	6.8 ±0.3
Arg62	6.67	6.57	6.62	6.62 ±0.03	_	_	_	_

Val63	6.44	6.29	6.31	6.35 ±0.05	5.72	5.56	5.06	5.5 ±0.2
Gly64	4.30	4.40	4.43	4.38 ±0.04	3.94	_	_	3.94

Table 2.4: $\Delta G^{o'}_{op}$, in kcal/mol, for the I57A;I37H variant (pH 6.5, 20 ^{o}C) in dilute solution. Protons that exchange upon global unfolding (82) are bolded. SEM, standard error of the mean.

Residue	Trial 1	Trial 2	Trial 3	Average ±SEM or range
Trp5	_	6.39	6.33	6.4 ±0.1
Leu8	5.96	5.89	5.82	5.89 ±0.04
Val9	5.65	5.77	5.53	5.7 ±0.1
Gly10	6.16	6.17	6.05	6.13 ±0.04
Lys11	**7.11**	**7.04**	**6.97**	**7.04 ±0.04**
Val13	6.15	6.16	6.03	6.11 ±0.04
Ala16	7.09	7.36	6.97	7.1 ±0.1
Lys17	6.70	6.66	6.53	6.6 ±0.1
Lys18	6.98	6.95	6.86	6.93 ±0.04
Val19	6.81	6.58	6.81	6.7 ±0.1
Ile20	**7.11**	**6.21**	**_**	**6.7 ±1.0**
Leu21	**_**	**6.91**	**6.06**	**6.5 ±0.9**
Gln22	6.86	6.86	6.74	6.82 ±0.04
Asp23	4.99	4.98	4.84	4.94 ±0.05
Lys24	7.27	7.15	7.14	7.19 ±0.04
Ala27	6.19	6.18	6.05	6.14 ±0.05
Gln28	6.00	5.99	_	6.00 ±0.01
Ile30	**7.39**	**6.75**	**_**	**7.1 ±0.6**
Val34	4.62	4.61	4.51	4.58 ±0.04
Arg46	6.68	6.67	6.58	6.64 ±0.03
Val47	**6.67**	**6.49**	**6.72**	**6.6 ±0.1**
Arg48	7.05	7.01	6.89	7.0 ±0.1
Leu49	**6.72**	**6.81**	**6.61**	**6.7 ±0.1**
Phe50	**6.66**	**6.50**	**6.45**	**6.5 ±0.1**
Val51	**6.58**	**7.06**	**6.71**	**6.8 ±0.1**
Asp52	6.36	6.25	6.18	6.3 ±0.1
Asp55	4.12	4.17	3.95	4.1 ±0.1
Asn56	7.06	7.05	6.93	7.01 ±0.04
Ala57	7.40	7.47	7.24	7.4 ±0.1
Ala58	6.92	6.95	6.86	6.91 ±0.03
Glu59	6.97	6.95	6.84	6.92 ±0.04
Arg62	7.26	7.34	7.21	7.3 ±0.04
Val63	6.14	6.19	6.04	6.12 ±0.04

Table 2.5: $\Delta G^{o'}_{op}$, in kcal/mol, for the I29A;I37H variant (pH 6.5, 20 ^{o}C) in 130.0 g dry weight /L, reconstituted cytosol. Protons that exchange upon global unfolding (82) are bolded. SEM, standard error of the mean.

Residue	Trial 1	Trial 2	Trial 3	Average ± SEM or range
Trp5	_	_	6.02	
Leu8	5.81	5.75	5.79	5.78 ±0.02
Val9	5.43	5.38	5.47	5.43 ±0.03
Lys11	**6.92**	**6.88**	**6.94**	**6.91 ±0.02**
Val13	5.94	5.91	6.02	5.96 ±0.03
Ala16	6.62	6.65	6.68	6.65 ±0.02
Lys18	5.93	5.98	6.09	6.0 ±0.1
Val19	6.54	_	_	_
Leu21	**6.19**	**_**	**_**	**_**
Gln22	6.40	6.42	6.40	6.41 ±0.01
Asp23	5.00	6.44	_	5.7 ±1.4
Lys24	6.42	6.36	6.33	6.37 ±0.03
Ala27	5.76	5.76	5.84	5.79 ±0.03
Val34	5.32	4.07	4.08	4.5 ±0.4
Arg46	6.48	6.45	6.51	6.48 ±0.02
Val47	**5.79**	**5.79**	**5.95**	**5.8 ±0.1**
Arg48	6.59	6.57	6.52	6.56 ±0.02
Leu49	**_**	**5.96**	**5.84**	**5.9 ±0.1**
Phe50	**6.09**	**5.96**	**5.98**	**6.01 ±0.04**
Val51	**5.94**	**5.99**	**_**	**6.0 ±0.1**
Asp52	6.15	6.05	6.05	6.08 ±0.03
Asp55	6.19	6.09	10.21	7.5 ±1.4
Asn56	5.95	5.95	11.34	7.8 ±1.8
Ala57	6.68	_	_	_
Ala58	6.51	6.35	6.41	6.42 ±0.05
Glu59	7.22	_	_	_
Arg62	6.42	_	6.42	6.42 ±0.00
Val63	6.14	6.06	6.04	6.08 ±0.03

Table 2.6: Assignments for the I57A;I37H variant (pH 6.5, 20 ^{o}C). na, not applicable.

Residue	^{1}H (ppm)	^{15}N (ppm)	Residue	^{1}H (ppm)	^{15}N (ppm)
Glu4	6.71	116.32	Gly35	9.09	115.80
Trp5	8.25	118.43	Thr36	7.58	117.55
Pro6	na	na	His37	8.91	127.01
Glu7	10.82	119.92	Val38	8.05	118.12
Leu8	7.93	118.72	Thr39	7.84	112.76
Val9	7.07	118.85	Met40	8.52	117.80
Gly10	9.15	116.14	Glu41	8.53	124.19
Lys11	7.95	119.04	Tyr42	8.65	125.42
Ser12	8.57	116.14	Arg43	8.75	132.03
Val13	8.33	120.62	Ile44	7.88	119.60
Glu14	8.32	117.52	Asp45	8.31	115.34
Glu15	7.66	120.39	Arg46	7.27	123.07
Ala16	8.88	120.30	Val47	7.98	122.73
Lys17	8.56	115.44	Arg48	9.05	126.43
Lys18	7.00	115.25	Leu49	8.90	126.17
Val19	7.40	119.24	Phe50	9.06	124.29
Ile20	8.03	119.47	Val51	8.85	115.13
Leu21	7.92	115.24	Asp52	8.75	123.82
Gln22	7.38	118.24	Lys53	8.26	115.96
Asp23	7.61	117.95	Leu54	8.04	121.31
Lys24	9.12	124.20	Asp55	8.13	115.02
Pro25	na	na	Asn56	8.06	113.97
Glu26	9.63	118.17	Ala57	9.13	122.64
Ala27	8.07	123.30	Ala58	9.63	128.77
Gln28	8.95	123.47	Glu59	7.36	114.09
Ile29	8.38	125.80	Val60	9.10	125.45
Ile30	8.08	130.30	Pro61	na	na
Val31	8.51	128.66	Arg62	8.56	122.24
Leu32	8.91	129.71	Val63	9.34	122.88
Pro33	na	na	Gly64	8.77	122.57
Val34	8.84	125.47			

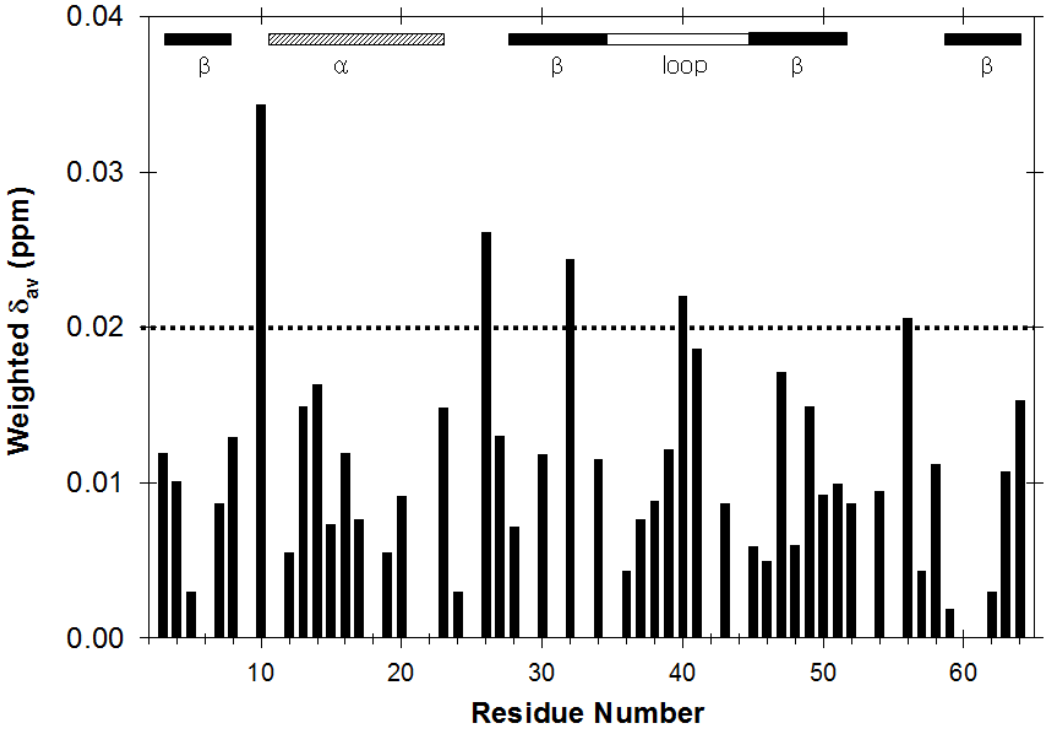

Figure 2.1: Small chemical shift changes show that the CI2 does not interact strongly with reconstituted cytosol. Two samples were used to calculate shift changes: 1 mM I29A;I37H variant in buffer and in 130.0 g dry weight /L of reconstituted cytosol (pH 6.5, 20 ^{o}C). Changes were calculated by using the equation (110):

$$\delta_{av} = [(\Delta^{1}\text{Hppm})^2 + \frac{(\Delta^{15}\text{Nppm x } 0.154)^2}{2}]^{1/2}$$

where δ_{av} is the shift in the reconstituted cytosol minus that in dilute solution. Values greater than 0.02 ppm are significant as shown from independent replicate experiments (84). The secondary structure is outlined above the histogram.

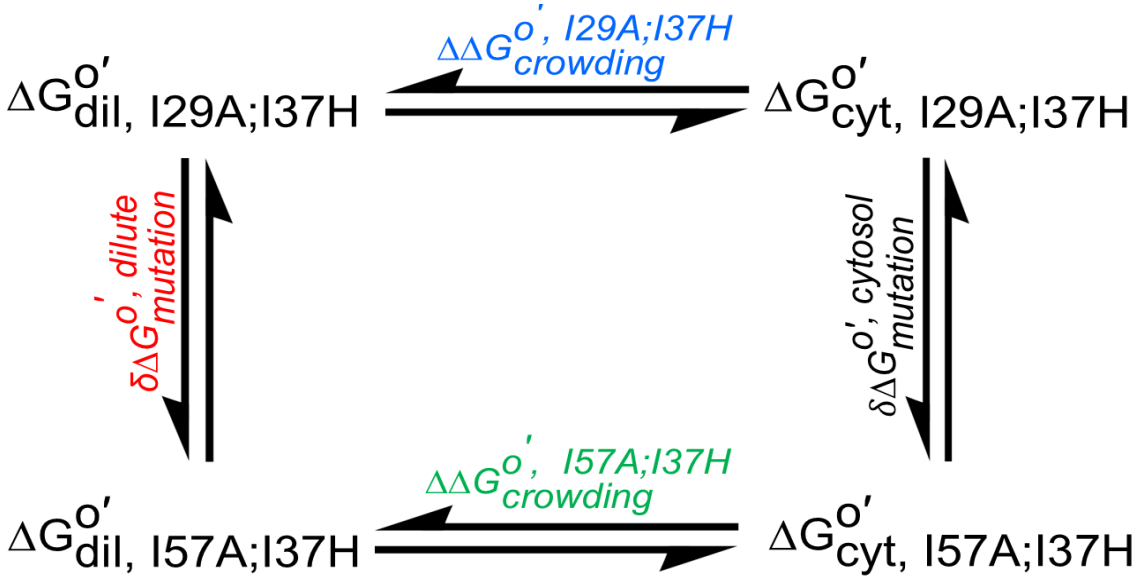

Figure 2.2 Thermodynamic cycle (cyt, cytosol; dil, dilute).

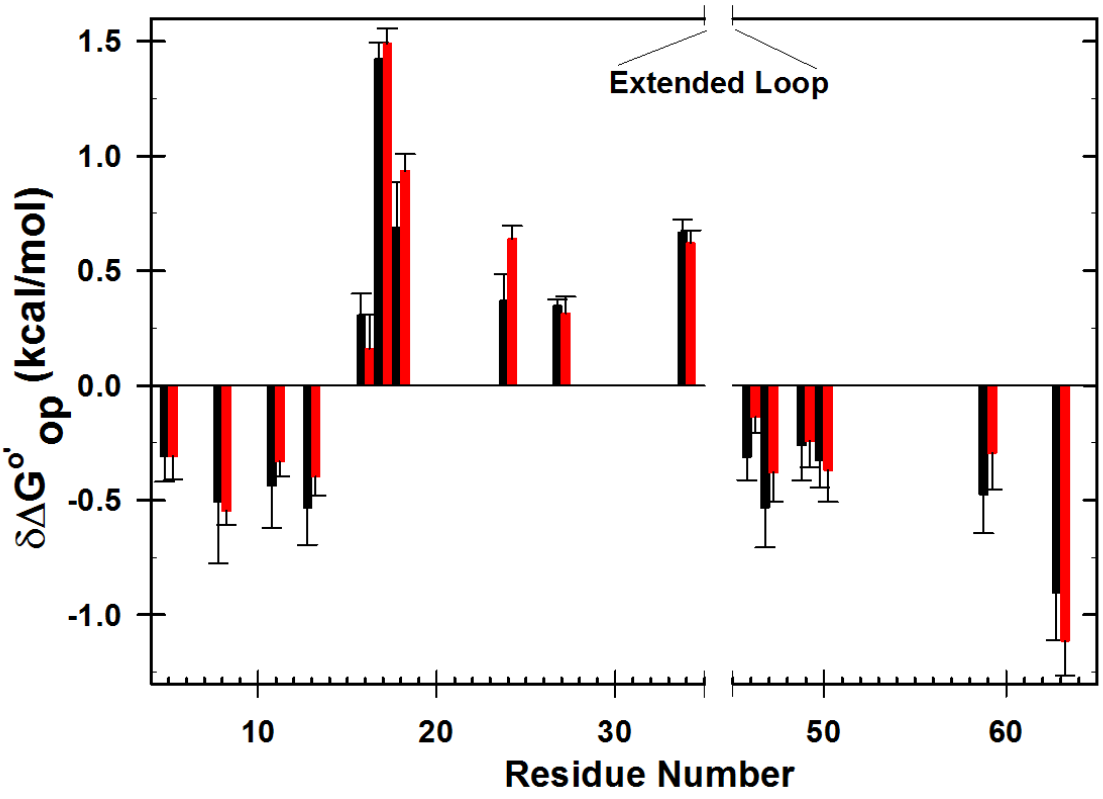

Figure 2.3 Differences in $\Delta G^{o'}_{op}$ corresponding to the vertical sides of Figure 2.2. Black bars, $\delta\Delta G^{o'}_{op}$ [= $\Delta G^{o'}_{op}$(I57A;I37H) - $\Delta G^{o'}_{op}$(I29A;I37H)] in reconstituted cytosol (100 g dry weight /L); Red bars, $\delta\Delta G^{o'}_{op}$ values in dilute solution. Data are shown for residues observed in both proteins whose $\delta\Delta G^{o'}_{op}$ values are statistically different from zero.

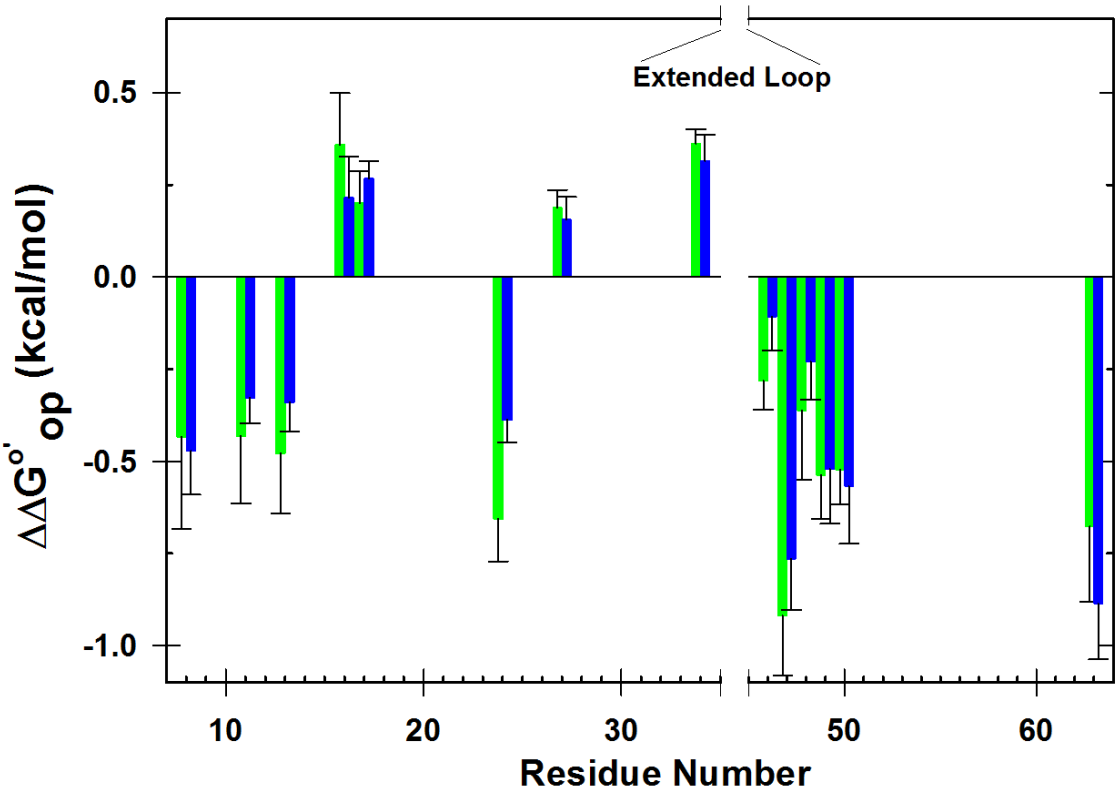

Figure 2.4. Differences in $\Delta G^{o'}_{op}$ corresponding to the horizontal sides of Fig. 1. Green bars, $\Delta\Delta G^{o'}_{op}$ [= $\Delta G^{o'}_{op}$(cytosol) - $\Delta G^{o'}_{op}$(dilute)] for I57A;I37H; blue bars, $\Delta\Delta G^{o'}_{op}$ values for I29A;I37H. Data are shown for residues observed in both proteins under both conditions whose $\Delta\Delta G^{o'}_{op}$ e values are statistically different from zero.

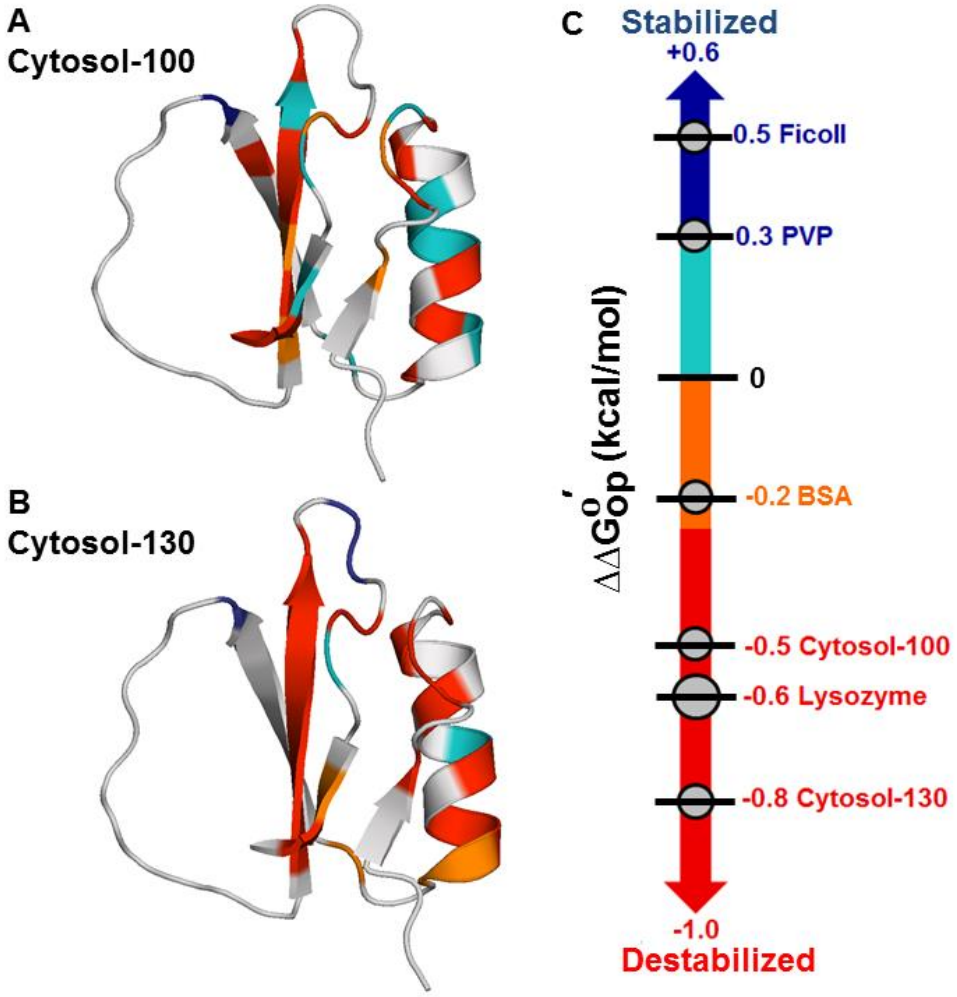

Figure 2.5. Impact of cytosol on stability. The backbone of CI2 (I29A;I37H) is color-coded by $\Delta\Delta G^{o'}_{op}$ in 100.0 g dry weight /L **(A)** and 130.0 g dry weight /L **(B)** of reconstituted cytosol [blue, $\Delta\Delta G^{o'}_{op}$ (kcal/mol) > 0.3; cyan, 0.3 ≥ $\Delta\Delta G^{o'}_{op}$ > 0.0; orange, 0.0 ≥ $\Delta\Delta G^{o'}_{op}$ > -0.3; red, $\Delta\Delta G^{o'}_{op}$ ≤ -0.3]. White areas denote prolines and amide protons that either exchanged completely before the second time point or were too broad to observe under one or more conditions. Changes in the average global stability of CI2 under crowded conditions **(C)**. The standard deviations of the mean of the $\Delta\Delta G^{o'}_{op}$ values for protons that exchange upon global

unfolding (86) are indicated by the size of the circle. Cytosol-100 and Cytosol-130 indicate 100.0 g dry weight /L and 130.0 g dry weight/L of the reconstituted cytosol, respectively at pH 6.5, 20 ^{o}C. Ficoll (100 g/L, pH 5.4, 37 ^{o}C) data are from (47). PVP (100 g/L, pH 5.4, 37 ^{o}C) data are from (43). BSA (100 g/L, pH 6.5, 20 ^{o}C) and lysozyme (100 g/L, pH 6.5, 20 ^{o}C) data are from (53). The lysate, BSA and lysozyme data were acquired at 20 ^{o}C. The stabilizing effects Ficoll and PVP are probably slightly exaggerated because these data were acquired at 37 ^{o}C, and crowding induced stability is predicted to increase with increasing temperature (99) .

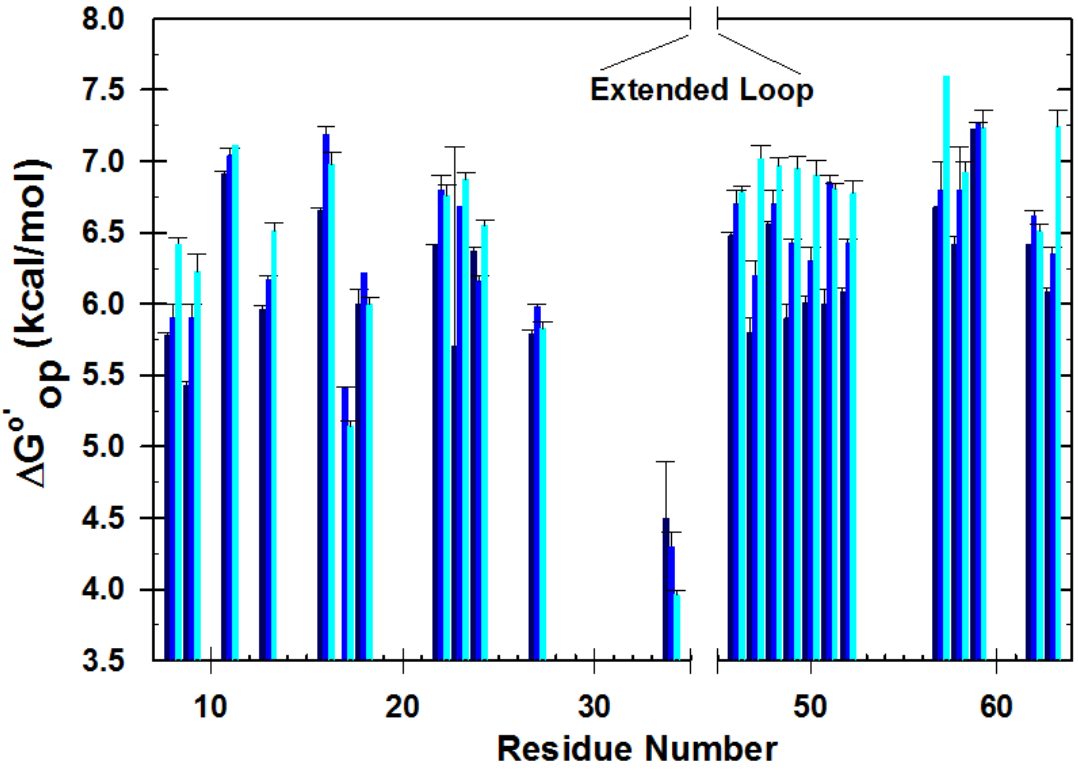

Figure 2.6: Residue level stability, $\Delta G^{o'}_{op}$ in kcal/mol, for the I29A;137H variant is plotted against residue number for three conditions: navy blue 130.0 g dry weight /L reconstituted cytosol; blue, 100.0 g dry weight /L reconstituted cytosol; cyan, buffer only (50 mM phosphate, pH 6.5, 20 ^{o}C for all conditions). The buffer data are from (53). The pattern is the same for crowded and dilute conditions, suggesting that CI2 folding is unaffected by the cytosol.

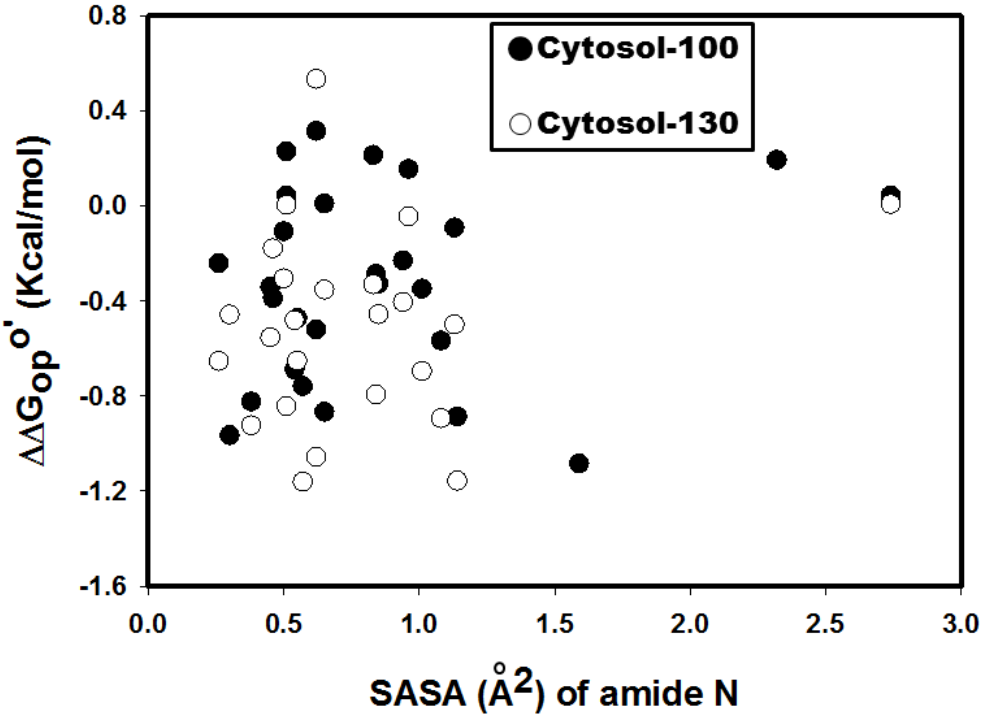

Figure 2.7: Changes in $\Delta G_{op}^{o'}$ *versus* solvent accessible surface areas (SASA) for 100.0 g dry weight /L and 130.0 g dry weight /L of reconstituted cytosol. Backbone amide SASA was calculated based on the crystal structure of CI2 (PDB ID: 2CI2) (111) using the online program POPS (112) with a 1.4 Å solvent probe.

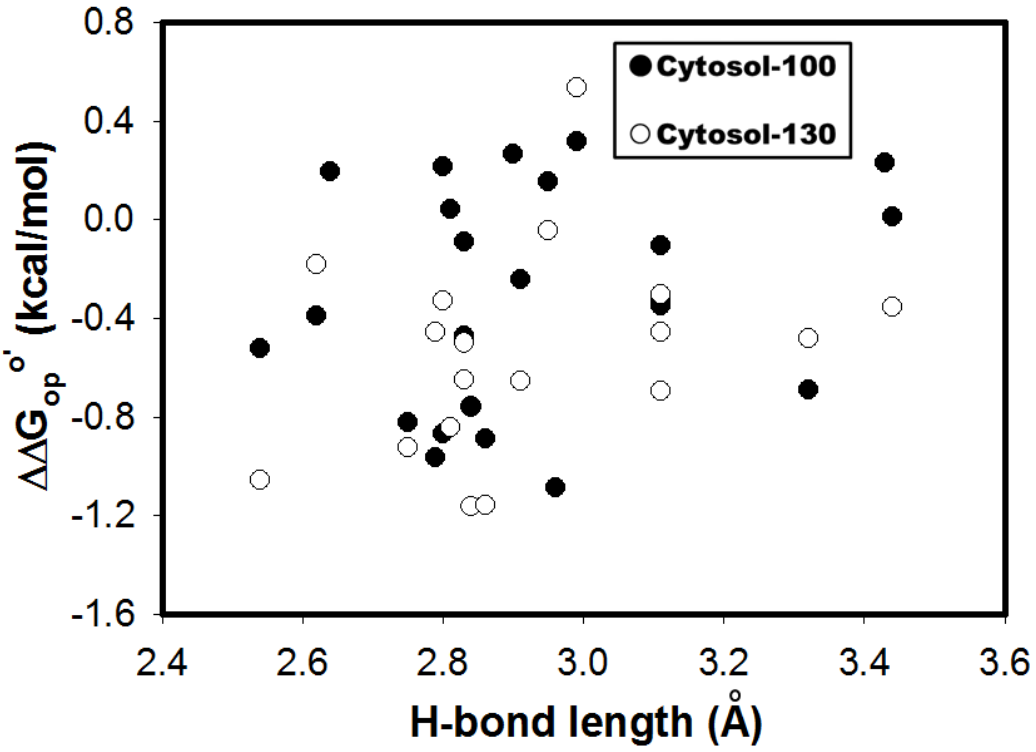

Figure 2.8: Changes in $\Delta G_{op}^{o'}$ *versus* intramolecular hydrogen bond lenghts in CI2 (111) for 100.0 g dry weight /L and 130.0 g dry weight /L of reconstituted cytosol calculated based on the crystal structure of CI2 (PDB ID: 2CI2) (111).

CHAPTER 3: PROTEIN-CROWDER CHARGE AND PROTEIN STABILITY

The material in this chapter is adapted from:

Sarkar M, Lu J and Pielak GJ. Protein-crowder charge and protein stability. *Manuscript submitted.*

(M.S. and G.J.P. designed research; M.S. and J.L. performed research; M.S. and G.J.P. analyzed data; M.S. and G.J.P. wrote the paper.)

3.1 Introduction

Traditionally, globular proteins were studied in dilute, buffered solutions even though their natural environment is crowded and heterogeneous (1). With the advent of sensitive biophysical techniques (78) and increased computer power (41, 54), biomolecules are now being studied in more physiologically relevant environments. Here, we focus on how the net charge of crowding molecules affects protein stability.

Globular protein stability is defined as the free energy of the biologically nonfunctional and less compact denatured state minus that of the functional and compact native state (2). Although crowding effects arise from both steric repulsions and weak nonspecific chemical interactions, only recently have these weak interactions been studied experimentally (16, 67, 70, 73). Steric and chemical interactions are often referred to as hard and soft, respectively (14). The hard component stabilizes a globular test protein because the decrease in space available to the test protein by the mere presence of the crowders favors compact species. The stabilization is entirely entropic because steric repulsions affect only the arrangement of solute molecules, not the chemical interactions between them.

Chemical interactions can be stabilizing or destabilizing. Soft repulsions reinforce the stabilizing influence of steric repulsions. On the other hand, destabilization of the test protein results when nonspecific attractive interactions dominate, because unfolding exposes attractive surface (53, 113). Nonspecific attractive interactions between proteins can also facilitate aggregation and misfolding. For example, macromolecular crowding fails to fold a marginally unstable variant of a globular protein (38). Therefore, it is important to study proteins under several conditions to assess the phenomena that give rise to crowding effects. Gaining an understanding of the nature of crowding will not only provide fundamental knowledge about biology, but also help solve practical problems. For instance, the knowledge will facilitate the design of synthetic polymers that increase the stability of industrially useful enzymes and protein based pharmaceuticals.

Crowding by synthetic polymers is often stabilizing (42, 43, 47), but crowding by individual proteins and by crude *Escherichia coli* lysates can destabilize globular proteins (53, 59, 113) and impede their diffusion (35) . Both effects arise from nonspecific attractive interactions (35, 114) . Biological crowders, such as bovine serum albumin (BSA), lysozyme and *E. coli* lysates, interact nonspecifically with the 7-kDa globular test protein chymotrypsin inhibitor 2 (CI2). The absence of significant CI2 chemical shift changes in unfractionated *E. coli* lysates shows that these interactions are weak, transient and nonspecific (113). Synthetic polymers are more inert (35). Furthermore, results from variable temperature studies of stability show that crowding effects can be primarily enthalpic (47, 58, 59), contrary to predictions based on the assumed primacy of steric repulsions.

CI2 has an isoelectric point (pI) of 6.0 has an estimated net charge of -1.0 at pH 7.0. We have studied the stability of CI2 when crowded by specific proteins (53). Having noted

that CI2 is more destabilized by lysozyme (pI 11.0), which has a net charge opposite to that of CI2, than by BSA (pI 4.7), which has the same net charge, we hypothesized that attractive charge-charge interactions overcome the stabilizing effect of steric interactions. The inside of cells, however, is not crowded by one particular protein. We studied CI2 in crude *E. coli* cell lysates and found that these lysates are also destabilizing (113). Here, we dissect the lysate to learn about the effects of net protein charge.

In our previous study we removed metabolites and other small molecules from the crude lysate by dialysis and showed that its protein component reflected that of the *E. coli* proteome. Nevertheless, the lysate still contained nucleic acids and nucleic-acid protein complexes (113). Here, we focus on the proteins in the lysate. Our hypothesis is that if the net charge of the test protein is same as the net charge of the crowding proteins, then charge-charge repulsions will dominate, enhancing the effects of steric repulsion, shifting the equilibrium towards the native state and stabilizing CI2.

To focus on the proteins in the lysate, we first removed the nucleic acids and their protein complexes with streptomycin sulfate and polyethylenimine (PEI) (115). Next, we used anion exchange chromatography at pH 7.0 in an attempt to divide the lysate proteins into two fractions, those with a net negative charge and those with a net positive charge. We then used NMR-detected amide proton exchange (78) at pH 7.0 to measure the stability of CI2 in buffer and in crowded solutions of the unfractionated and fractionated lysates.

3.2 Results

3.2.1 Lysate Characterization

Previous analyses show that our dialyzed *E. coli* lysate contains proteins with pI values from 4 to 13 (113), similar to the range predicted from inspecting the *E. coli* proteome (37), and that a 100 g dry weight/L lysate solution contained 52 ±4 g of protein/L, with remainder comprising mostly nucleic acids (74). We prepared a refined dialyzed lysate depleted in nucleic acids and nucleic acid binding proteins *via* streptomycin and PEI precipitations (115). We call this the total protein lysate. A 100 g dry weight/L solution of this lysate contained 92 ±4 g of protein/L, attesting to the effectiveness of the precipitations.

We attempted to divide the total protein lysate into cationic and anionic fractions by using anion exchange chromatography at pH 7.0. Proteins with pI values less than 7.0 should be retained, and those with values greater than 7.0 should be eluted by low salt buffer. The bound proteins were eluted with high salt buffer and dialyzed. The two fractions were lyophilized and analyzed by mass spectrometry. Two hundred proteins were identified in the bound fraction and 193 proteins were identified in the flow through (Tables 3.1 and 3.2 of the Supporting Information). Only six of the bound proteins had a pI value greater than 7.0, affirming the efficiency of chromatography. We define this sample as the anionic protein lysate. A 100 g dry weight/L solution of this lysate contained 95 ±2 g of protein/L. In summary, the anionic lysate comprised almost exclusively anionic proteins at pH 7.0.

Ideally, the proteins with pI values greater than 7.0 should flow through an anion exchange column. Analysis revealed, however, that this fraction comprised proteins with pI values ranging from 4.4 to 11.4. Control experiments showed that the column was not overloaded. Of the 193 proteins identified, only 73 possessed a pI greater than 7.0.

Furthermore, twelve of the 193 proteins were also identified in the bound fraction (highlighted in yellow in Tables 3.1 and 3.2 of the Supporting Information). We did not use this fraction because it could not be classified as either anionic or cationic.

The distribution of proteins with respect to pI was assessed (Figure 1). We summed the lists of bound and unbound proteins to give the distribution in the total protein lysate. The inset in Figure 1 shows the distribution from analyzing the *E. coli* genome (37). Both histograms exhibit similar shapes, indicating that the total protein lysate is a reasonable approximation of the proteome.

3.2.2 Stability Effects

We assessed amide ^{1}H exchange of 1 mM CI2 in buffer and in 100.0-g dry weight/L solutions of the total protein lysate and anionic protein lysate at pH 7.0 and 20 ^{o}C in triplicate by using NMR (78). The observed rates of exchange (k_{obs}, s^{-1}) can be converted to free energies of opening ($\Delta G_{op}^{o'}$) if the intrinsic rates of exchange (k_{int}, s^{-1}) are rate determining and unchanged by crowding (77, 78) These conditions are satisfied for lysates at pH 6.5, 20 ^{o}C (79, 113). An increase of pH from 6.5 to 7.0 does not change the stability of CI2 (Table 3.1).

$\Delta G_{op}^{o'}$ values and their uncertainties (Tables 3.3, 3.4 and 3.5) were used to calculate stability differences, $\Delta\Delta G_{op}^{o'}$ (crowded - buffer), and the propagated standard error.(116) Both lysates result almost exclusively in $\Delta\Delta G_{op}^{o'}$ of less than zero, indicating that the lysates destabilize CI2 (Figure 2).

Destabilization was observed along the entire backbone (Figures 2 and 3), except at Val 34. The unstructured loop of CI2 begins immediately after this residue. Although exchange is too fast to quantify loop stability with the method used here, chemical exchange based experiments indicate that crowding does not affect exchange rates in this region (79).

Crude lysate also stabilizes Val 34, and stabilization increases with lysate concentration (113). Perhaps one or more proteins block exchange by interacting specifically with the loop, or repulsive interactions alter its conformation.

Global stabilities (Table 3.6) were calculated by averaging the $\Delta G_{op}^{o'}$ values for globally exchanging residues (82). Contrary to our hypothesis, both the total protein lysate and the anionic protein lysate destabilized CI2, and by the same amounts, -0.6 ± 0.1 kcal/mol.

3.3 Discussions

3.3.1 p*I Bias of the* Proteome

Many proteomes exhibit a bimodal distribution of isoelectric points (117). The bias in *E. coli* is toward anionic proteins (Figure 1) (37). Not only is the number of acidic proteins greater, so is their abundance (41, 83). This observation explains why it was easier for us to extract the anionic proteins. Many of the cationic proteins were probably lost during the streptomycin sulfate- or PEI- precipitation because of their interactions with nucleic acids, which might explain why we were unable to obtain a useful cationic protein lysate. Others have observed this bias against isolating basic proteins from *E. coli* genome (83).

3.3.2 Anionic Protein Lysate has Same Effect as Total Protein Lysate

It is intuitive that anionic proteins will repulse negatively charged CI2. This repulsion should reinforce crowding-induced steric repulsion, favoring the compact native state over the ensemble of larger unfolded states, and thus increase stability. However, we observed only destabilization. Importantly, the anionic protein lysate is nearly as destabilizing as total protein lysate (Table 3.1). This observation is consistent with the idea

that nonspecific attractive backbone interactions shift the equilibrium towards the less structured states.

A negatively charged surface should increase repulsion, yet we observe uniform destabilization. Furthermore, we find no correlation between CI2 surface potential and residue level destabilization (Figure 3). We conclude that neither the net charge of CI2 nor the net charges of the crowding proteins can overcome the nonspecific attractive interactions between protein surfaces. This conclusion is consistent with one of our observations on the destabilizing effect of crowding by BSA (53). Specifically, adding NaCl to screen the charge mitigated, but did not eliminate, destabilizing interactions.

3.4 Conclusions

We have studied crowding effects on the stability of CI2 under several conditions (42, 43, 47, 53, 113). Synthetic polymers such as PVP and Ficoll are stabilizing (42, 43, 47), but all the physiologically relevant crowders studied so far destabilize CI2 (53, 113), Protein L (38) and ubiquitin near room temperature (59). We also know that CI2 interacts more strongly with biologically relevant crowders than with these synthetic polymers (35, 114). Even the anionic proteins, which have the same net charge as CI2, interact strongly enough with the backbone to overcome both charge-charge and steric repulsions. We conclude that proteins possess an inherently favorable, and probably ubiquitous, interaction with other proteins. Although weak, these interactions can overcome the stabilizing effect of hardcore repulsions associated with physiologically relevant macromolecular crowding.

3.5 Materials and Methods

3.5.1 Chymotrypsin inhibitor 2 (CI2)

^{15}N enriched I29A;I37H CI2 was expressed in M9 media and purified as described (42, 43, 53, 79).

3.5.2 *Escherichia coli* Extracts (79)

Typically, six, 1-L cultures of *E. coli* were grown to saturation in Luria Broth (LB) containing kanamycin (60.0 µg/mL). The cells were pelleted (6500 *g*, 30 min) and the pellets stored at -20 ^{o}C after adding protease inhibitors (Sigma-Aldrich; 0.02 mM 4-(2-aminoethyl) benzenesulfonyl fluoride, 0.14mM E-64, 1.30 mM bestatin, 0.01 mM leupeptin, 3.0 nM aprotinin and 0.01 mM sodium EDTA, final concentrations). Each pellet was resuspended in 25.0 mL of Buffer A (50 mM sodium phosphate, pH 7.0) and lysed by sonic dismembration on ice for 8 min (Fisher Scientific, Sonic Dismembrator Model 500, 20% amplitude, 2 s on, 3 s off). The insoluble debris was removed by centrifugation (14,000 *g*, 10 ^{o}C, 30 min). Approximately 180 mL of supernatant was obtained.

3.5.3 Total Protein Lysate

Streptomycin sulfate and polyethylenimine (PEI) were used to deplete the lysate of nuclei acids and protein-nucleic acid complexes (115). Streptomycin sulfate (1.8 g, 10.0 mg/mL final concentration) was added, and the sample was stirred on ice for 30 min. The solution was centrifuged at 14,000 *g*, at 10 ^{o}C for 30 min, and the supernatant retained. PEI was added to a final concentration of 0.1% (w/v) over the course of 1 h while stirring on ice, and the sample centrifuged as described above. The supernatant was retained. PEI precipitation was repeated until the ratio of the absorbance at 260 nm to the absorbance at 280 nm of the supernatant was <1.2. The supernatant was dialyzed (Thermo Scientific,

SnakeSkin, 3K MWCO) against 5 L of 10 mM sodium phosphate buffer (pH 7.0) at 4 ^{o}C for 72 h. The dialysate was centrifuged at 14,000 *g*, 10 ^{o}C for 1 h. The insoluble nucleic acid- and protein nucleic acid- complexes were discarded. The sample was sterile filtered (0.22 µm Durapore® PVDF membrane, Millipore Corporation) and protease inhibitors were added as described above. The sample was lyophilized (Labconco) and stored at -20 ^{o}C.

3.5.4 Anionic Protein Lysate

Bench-top anion exchange chromatography was performed at room temperature. A 20 cm by 50 mm inner diameter column was packed with a 450-mL slurry of diethylaminoethyl (DEAE) cellulose (GE Healthcare). The final column volume was ~250 mL. Column was equilibrated at room temperature with 500 mL of Buffer A at a flow rate of 1.0 mL/min, and 180 mL of total protein fraction was loaded. The column was washed (1.0 mL/min) with 1.0 L of Buffer A and the flow through was retained. Bound proteins were eluted at 1.0 mL/min with 500 mL of Buffer B (Buffer A plus 1 M NaCl, pH 7.0). Both fractions were lyophilized after adding protease inhibitors. The bound fraction was resuspended in 100 mL of Buffer A and dialyzed (Thermo Scientific, SnakeSkin, 3K MWCO) against 5 L of 10 mM sodium phosphate buffer, pH 7.0 for 48 h, changing the buffer every 24 h. The sample was then lyophilized to a straw colored powder and stored at -20 ^{o}C.

3.5.5 Mass Spectrometry

Proteomic analyses were performed at the UNC Michael Hooker Proteomics Center. Raw MS files corresponding to peptide fragments from each mixture were searched by using MASCOT software (Matrix Science, Ver 2.3.02) *via* Proteome Discoverer (Thermo, Ver. 1.3.0.339) against the UniProt *E. coli* database (109). Nearly all the bound proteins had

isoelectric points (p*I*) of less than 7.0 (Table 3.1). We refer to the bound proteins fraction as 'anionic proteins' because experiments were performed at pH 7.0. The unbound fraction had p*I* values ranging from 4.0 to 11.5 (Table 3.2). We did not use this fraction.

3.5.6 Amide ^{1}H Exchange

Experiments were performed in 100% deuterated solutions (50 mM sodium phosphate, pH 7.0) at 20 ^{o}C unless stated otherwise. Labile protons in the lysates were pre-exchanged with deuterons as described (78).Three solutions were prepared. The first solution comprised Buffer A. The second solution was made by suspending 600.0 mg of total protein lysate in Buffer A to a volume of 6.0 mL. The third solution was prepared by suspending 500.0 mg of the anionic protein lysate in Buffer A to a volume of 5.0 mL. The pH_{read} of all solutions was adjusted to 7.0 (108). Lysate containing samples were centrifuged at room temperature at 14,000 *g* for 10 min to remove insoluble material. A modified Lowry assay (Thermo Scientific) was performed on each supernatant after a 100-fold dilution . The 100 g/L of total protein solution contained 92 ±4 g of protein/L of proteins. The 100 g/L anionic protein solution contained 95 ±2 g of protein/L.

Triplicate NMR-detected amide ^{1}H exchange experiments were performed on 1 mM CI2 samples (800 µl in a 5-mm Norell tube,) in Buffer A, 100 g/L total protein lysates and 100 g/L anionic protein lysate. Each experiment required 24 h. Sample preparation, shimming and acquisition of serial heteronuclear single quantum coherence (HSQC) spectra have been described (53, 78). Briefly, lyophilized ^{15}N enriched CI2 was added to 1.0 mL of Buffer A. The CI2 concentration was determined measuring the absorbance at 280 nm [ε = 7.04 X 10^{3} $M^{-1}cm^{-1}$] (118) The same mass of CI2 was added to lysate solutions, immediately prior to starting acquisition. The shims of the 600 MHz spectrometer (Varian Inova, room

temperature, triple resonance, HCN probe) were optimized with a 1 mM CI2 sample in Buffer A (10% v/v D_2O) . Twenty-two serial HSQC spectra were acquired with 1024 complex points in the ^{1}H dimension and 64 complex increments in the ^{15}N dimension. Intrinsic rates *(k_{int})* were calculated with SPHERE (119) at pH_{read} 7.0, 20 ^{o}C, 100% D_2O. Data were processed with NMRPipe (106) and NMRView (107).

3.6 Tables and Figures

Table 3.1 Proteomic analysis of proteins that bound the diethylaminoethyl (DEAE) cellulose column. The list comprises proteins with two or more identified peptides, their percent sequence coverage and calculated isoelectric point (p*I*). Of the 200 proteins, 194 had p*I* values of less than 7.0. Proteins that are also identified in the unbound fraction are highlighted in yellow.

Peptides	%Coverage	Names	p*I*
32	56.02	60 kDa chaperonin	4.9
25	64.21	Elongation factor Tu 1	5.5
24	40.91	Elongation factor G	5.4
23	60.51	Tryptophanase	6.2
21	58.99	Isocitrate lyase	5.3
19	35.74	Chaperone protein DnaK	5.0
18	57.88	Phosphoglycerate kinase	5.2
18	58.51	Succinyl-CoA ligase [ADP-forming] subunit beta	5.5
18	40.95	Acetyl-coenzyme A synthetase	5.8
17	26.49	Aconitate hydratase 1	5.9
15	22.20	Aconitate hydratase 2	5.4
15	45.90	Citrate synthase	6.7
14	45.91	Isocitrate dehydrogenase [NADP]	5.3
13	50.21	Aspartate ammonia-lyase	5.3
13	58.01	Glyceraldehyde-3-phosphate dehydrogenase A	7.1
12	39.12	Enolase	5.5
11	36.11	Phosphoenolpyruvate carboxykinase [ATP]	5.7
11	30.64	Galactitol-1-phosphate 5-dehydrogenase	6.4
10	52.68	Transaldolase B	5.2
10	17.97	Chaperone protein ClpB	5.5
10	31.22	Transketolase 1	5.7
10	15.67	Glycine dehydrogenase [decarboxylating]	6.0
10	23.19	Pyruvate kinase I	6.1
9	15.75	Carbamoyl-phosphate synthase large chain	5.3
9	28.29	Phosphate acetyltransferase	5.4
9	22.32	Pyruvate dehydrogenase E1 component	5.7
9	37.14	D-tagatose-1,6-bisphosphate aldolase subunit GatZ	5.8
9	35.61	Aspartate aminotransferase	5.8
9	36.49	Fructose-bisphosphate aldolase class 2	5.9
9	65.15	Flavoprotein WrbA	5.9
9	33.14	Fructose-bisphosphate aldolase class 1	6.7

8	30.22	Phosphopentomutase	5.3
8	36.69	Argininosuccinate synthase	5.4
8	27.85	Dihydrolipoyl dehydrogenase	6.2
8	30.99	4-aminobutyrate aminotransferase GabT	6.2
8	35.96	Succinylornithine transaminase	6.4
8	28.78	Serine hydroxymethyltransferase	6.5
8	25.00	Fumarate hydratase class I, aerobic	6.6
8	39.79	Succinyl-CoA ligase [ADP-forming] subunit alpha	6.8
7	25.82	Glutamate-1-semialdehyde 2,1-aminomutase	4.8
7	12.12	Phosphoenolpyruvate synthase	5.1
7	45.99	Alkyl hydroperoxide reductase subunit C	5.2
7	41.22	Enoyl-[acyl-carrier-protein] reductase [NADH] FabI	5.9
7	26.18	Soluble pyridine nucleotide transhydrogenase	6.6
7	45.49	Glucosamine-6-phosphate deaminase	6.9
6	43.79	Glucose-specific phosphotransferase enzyme IIA component	4.8
6	21.28	DNA-directed RNA polymerase subunit alpha	5.1
6	21.09	Lactaldehyde dehydrogenase	5.2
6	49.48	10 kDa chaperonin	5.2
6	27.92	Elongation factor Ts	5.3
6	15.88	Asparagine--tRNA ligase	5.3
6	13.31	NADP-dependent malic enzyme	5.5
6	20.44	3-oxoacyl-[acyl-carrier-protein] synthase 1	5.5
6	30.77	Aminomethyltransferase	5.6
6	16.33	CTP synthase	5.9
6	34.67	Cysteine synthase A	6.1
6	45.60	2,3-bisphosphoglycerate-dependent phosphoglycerate mutase	6.2
6	24.26	Cysteine desulfurase	6.4
6	10.33	Aldehyde-alcohol dehydrogenase	6.8
5	18.98	Trigger factor	4.9
5	22.26	ADP-L-glycero-D-manno-heptose-6-epimerase	4.9
5	4.47	DNA-directed RNA polymerase subunit beta	5.3
5	20.39	6-phosphofructokinase isozyme 2	5.4
5	8.57	Glycerol kinase	5.5
5	19.07	Aspartate-semialdehyde dehydrogenase	5.6
5	20.31	Aldehyde dehydrogenase B	5.6
5	48.95	Nucleoside diphosphate kinase	5.8
5	19.44	Peptidase B	5.9
5	19.10	2-amino-3-ketobutyrate coenzyme A ligase	6.0
5	16.46	Gamma-aminobutyraldehyde dehydrogenase	6.0
5	10.26	Formate acetyltransferase 1	6.0
5	14.12	Succinate dehydrogenase flavoprotein subunit	6.3
5	20.00	2,5-diketo-D-gluconic acid reductase A	6.5
5	21.80	Arginine N-succinyltransferase	6.5
5	49.11	UPF0438 protein YifE	6.6

5	16.67	Pyruvate kinase II	6.7
5	5.33	DNA-directed RNA polymerase subunit beta'	7.1
4	36.52	Glutaredoxin-4	4.8
4	43.67	Bacterioferritin	4.8
4	15.22	ATP synthase subunit beta	5.0
4	9.19	6-phosphogluconate dehydrogenase, decarboxylating	5.2
4	8.74	Proline--tRNA ligase	5.2
4	31.58	S-ribosylhomocysteine lyase	5.4
4	15.77	Succinate-semialdehyde dehydrogenase [NADP(+)] GabD	5.6
4	10.99	Uncharacterized ABC transporter ATP-binding protein YjjK	5.6
4	21.76	Purine nucleoside phosphorylase DeoD-type	5.7
4	31.76	Triosephosphate isomerase	6.0
4	4.90	Isoleucine--tRNA ligase	6.0
4	6.96	Ribonucleoside-diphosphate reductase 1 subunit alpha	6.2
4	19.48	Uncharacterized zinc-type alcohol dehydrogenase-like protein YahK	6.2
4	10.14	Pyruvate dehydrogenase	6.3
4	13.49	Transketolase 2	6.3
4	14.66	L-threonine 3-dehydrogenase	6.4
3	39.29	Thiol peroxidase	4.9
3	9.03	tRNA modification GTPase MnmE	5.0
3	32.05	Putative peroxiredoxin bcp	5.2
3	9.41	Oligopeptidase A	5.3
3	5.63	Aminopeptidase N]	5.3
3	23.14	7-alpha-hydroxysteroid dehydrogenase	5.4
3	18.27	Uracil phosphoribosyltransferase	5.5
3	13.81	Phosphoserine aminotransferase	5.6
3	8.44	Protein pmbA	5.6
3	13.55	Phosphoglucomutase	5.7
3	17.81	6-phosphofructokinase isozyme 1	5.7
3	15.69	2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltransferase	5.7
3	9.43	Tyrosine--tRNA ligase	5.8
3	18.26	NADPH-dependent curcumin reductase	5.8
3	9.03	Glucosamine--fructose-6-phosphate aminotransferase [isomerizing]	5.9
3	6.51	Catalase HPII	5.9
3	10.81	Putative glyceraldehyde-3-phosphate dehydrogenase C	6.1
3	16.46	3-oxoacyl-[acyl-carrier-protein] synthase 2	6.1
3	30.41	Ferric uptake regulation protein	6.1
3	8.50	Fatty acid oxidation complex subunit alpha	6.2
3	11.19	Glutamine--tRNA ligase	6.3
3	12.57	Carbamoyl-phosphate synthase small chain	6.4

3	24.43	Putative UPF0012 hydrolase YbeM	6.4
3	16.60	Inosine-5'-monophosphate dehydrogenase	6.4
3	5.36	2-oxoglutarate dehydrogenase E1 component	6.5
3	16.92	Fumarate hydratase class II	6.6
3	8.33	Bifunctional polymyxin resistance protein ArnA	6.9
3	21.31	3-oxoacyl-[acyl-carrier-protein] reductase FabG	7.4
3	26.06	50S ribosomal protein L10 unit	9.0
2	18.71	Protein-export protein SecB	4.4
2	7.42	USG-1 protein	4.5
2	11.11	Transcription elongation protein NusA	4.6
2	21.49	50S ribosomal protein L7/L12 unit	4.7
2	16.37	3-mercaptopyruvate sulfurtransferase	4.7
2	4.87	Phosphoenolpyruvate-protein phosphotransferase	4.9
2	18.58	Vitamin B12 transport periplasmic protein BtuE	4.9
2	25.78	NifU-like protein	4.9
2	19.68	Glutaminase 1	5.0
2	10.96	PTS-dependent dihydroxyacetone kinase, dihydroxyacetone-binding subunit dhaK	5.0
2	13.20	Glycine--tRNA ligase alpha subunit	5.0
2	13.92	2-dehydro-3-deoxygluconokinase	5.0
2	24.19	UPF0502 protein YceH	5.1
2	11.44	Molybdopterin molybdenumtransferase	5.1
2	4.29	Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex	5.2
2	8.79	D-alanine--D-alanine ligase A	5.2
2	14.17	UPF0135 protein YbgI	5.2
2	9.94	Chaperone protein HtpG	5.2
2	6.14	Lysine--tRNA ligase, heat inducible	5.2
2	31.41	6,7-dimethyl-8-ribityllumazine synthase	5.3
2	7.29	S-adenosylmethionine synthase	5.3
2	4.60	Argininosuccinate lyase	5.3
2	25.00	Universal stress protein A	5.3
2	4.53	Phenylalanine--tRNA ligase beta subunit	5.3
2	5.00	Leucine--tRNA ligase	5.3
2	6.55	NAD-dependent malic enzyme	5.3
2	7.33	Ketol-acid reductoisomerase	5.3
2	13.45	Aerobic respiration control protein ArcA	5.3
2	19.34	Stringent starvation protein A	5.3
2	4.31	Valine--tRNA ligase	5.3
2	8.35	ATP-dependent protease ATPase subunit HslU	5.4
2	7.42	Cytosol non-specific dipeptidase	5.4
2	16.89	Ribose-5-phosphate isomerase A	5.4
2	5.87	Bifunctional protein HldE	5.4
2	7.26	Glycine--tRNA ligase beta subunit	5.4
2	10.95	Protein yeeZ	5.5
2	8.96	Glutamine synthetase	5.5

2	5.37	Arginine--tRNA ligase	5.5
2	5.58	Serine--tRNA ligase	5.5
2	10.90	Thioredoxin reductase	5.5
2	9.03	ATP phosphoribosyltransferase	5.6
2	10.78	Ornithine carbamoyltransferase chain I	5.7
2	6.72	Alkyl hydroperoxide reductase subunit F	5.7
2	4.70	Peptidyl-dipeptidase dcp	5.7
2	8.33	Malate dehydrogenase	5.8
2	14.81	Glyoxylate/hydroxypyruvate reductase B	5.8
2	6.87	Alpha-galactosidase	5.8
2	3.25	Methionine--tRNA ligase	5.9
2	7.04	Asparagine synthetase B [glutamine-hydrolyzing]	5.9
2	6.78	UDP-N-acetylmuramate:L-alanyl-gamma-D-glutamyl-meso-diaminopimelate ligase	5.9
2	5.03	Dihydroxy-acid dehydratase	5.9
2	7.43	Periplasmic trehalase	5.9
2	6.37	Histidine--tRNA ligase	5.9
2	4.65	5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase	5.9
2	2.05	Bifunctional protein putA	6.0
2	9.48	Xaa-Pro dipeptidase	6.0
2	7.53	Fructose-1,6-bisphosphatase class 1	6.0
2	13.31	NADP-dependent 3-hydroxy acid dehydrogenase YdfG	6.0
2	9.89	Molecular chaperone Hsp31 and glyoxalase 3	6.0
2	8.92	Protein CsiD	6.2
2	16.21	Uridine phosphorylase	6.2
2	9.85	Acetylornithine/succinyldiaminopimelate aminotransferase	6.2
2	12.90	Oxygen-insensitive NAD(P)H nitroreductase	6.2
2	46.81	Galactitol-specific phosphotransferase enzyme IIB component	6.3
2	12.66	Transaldolase A	6.3
2	9.00	Acetate kinase	6.3
2	11.65	Uncharacterized protein YciO	6.4
2	16.59	Thymidine kinase	6.4
2	6.94	NADP-specific glutamate dehydrogenase	6.4
2	16.67	Phosphoheptose isomerase	6.4
2	26.85	50S ribosomal protein L9 unit	6.6
2	3.72	2,4-dienoyl-CoA reductase [NADPH]	6.6
2	15.93	Ribosome-associated inhibitor A	6.7
2	11.56	Protein tas	6.8
2	21.31	Putative NAD(P)H nitroreductase ydjA	6.8
2	13.70	3-ketoacyl-CoA thiolase	6.8
2	14.44	2-dehydro-3-deoxyphosphooctonate aldolase	6.8
2	20.40	Osmotically-inducible protein Y	6.8

2	28.17	Universal stress protein D	6.9
2	16.22	GTP cyclohydrolase 1	7.3
2	9.96	Uridylate kinase	7.9

Table 3.2 Proteomic analysis of proteins that did not bind to the DEAE cellulose column. The list comprises proteins with two or more identified peptides, their percent sequence coverage and calculated p*I*. Of the 193 proteins identified, only 71 had p*I* greater than 7.0. Proteins that are also identified in the bound fractions are highlighted in yellow.

Peptides	**%Coverage**	**Names**	**p*I***
37	94.23	Malate dehydrogenase	5.8
34	79.15	Glyceraldehyde-3-phosphate dehydrogenase A	7.1
22	57.08	sn-glycerol-3-phosphate-binding periplasmic protein	6.7
21	71.92	Lysine-arginine-ornithine-binding periplasmic protein	5.9
21	71.2	Transaldolase A	6.3
21	71.19	Glutamate/aspartate periplasmic-binding protein	8.5
20	57.27	Periplasmic oligopeptide-binding protein	6.5
20	59.63	Periplasmic dipeptide transport protein	6.7
19	78.11	Osmotically-inducible protein Y	6.8
17	68.85	Histidine-binding periplasmic protein	5.6
16	33.08	2',3'-cyclic-nucleotide 2'-phosphodiesterase/3'-nucleotidase	5.7
15	70.78	Putative ABC transporter arginine-binding protein 2	6.1
15	62.79	Glutaredoxin-2	8.0
14	58.14	Protein yhbO	5.5
14	32.64	Malate synthase G	6.2
14	59.12	ABC transporter periplasmic-binding protein YtfQ	7.1
14	55.91	Putative ABC transporter periplasmic-binding protein YdcS	7.4
13	63.55	Adenylate kinase	5.8
13	30.29	Biotin carboxylase	7.1
13	59.32	50S ribosomal protein L6	9.7
12	60	Ribosome-recycling factor	6.9
11	95.65	UPF0337 protein YjbJ	5.6
11	58.04	Peroxiredoxin OsmC	5.9
11	50.34	D-ribose-binding periplasmic protein	7.5
11	45.56	Glutamine-binding periplasmic protein	8.5
10	91.19	Superoxide dismutase [Fe]	5.9
10	69.23	Uncharacterized protein YahO	6.1
10	28.9	Outer membrane protein A	6.4
10	52.73	2,5-diketo-D-gluconic acid reductase A	6.5
10	35.19	Erythronate-4-phosphate dehydrogenase	6.7
10	48.32	Quinone oxidoreductase 1	8.3
10	47.41	FKBP-type peptidyl-prolyl cis-trans isomerase FkpA	8.5
9	61.35	UPF0234 protein YajQ	6.2
9	76.21	Superoxide dismutase [Mn]	7.0
8	78.91	Enamine/imine deaminase	5.5

8	51.64	KHG/KDPG aldolase	5.7
8	41.46	3'(2'),5'-bisphosphate nucleotidase CysQ	6.0
8	25	Uncharacterized sulfatase YdeN	6.0
8	15.56	Periplasmic beta-glucosidase	6.2
8	48.98	Uncharacterized oxidoreductase YghA	6.8
8	41.98	ABC transporter arginine-binding protein 1	7.4
8	29.77	Protein TolB	7.6
7	61.47	Thioredoxin-1	4.9
7	40.15	Deoxyribose-phosphate aldolase	5.7
7	46.95	Peptidyl-prolyl cis-trans isomerase B	5.8
7	40.53	Methionine aminopeptidase	6.0
7	24	Uncharacterized protein YjgR	6.3
7	21.9	Probable D,D-dipeptide-binding periplasmic protein DdpA	6.7
7	59.12	Transcription antitermination protein NusG	6.8
7	41.98	Ecotin	7.1
7	39.49	Protein YdgH	9.3
7	58.45	50S ribosomal protein L11	9.6
6	79.52	Protein YciN	5.7
6	85.88	Phosphocarrier protein HPr	5.8
6	34.32	Putative carboxymethylenebutenolidase	6.0
6	45.37	Nitrate/nitrite response regulator protein NarL	6.1
6	15.25	Glucans biosynthesis protein D	6.3
6	64.74	Superoxide dismutase [Cu-Zn]	6.4
6	64.63	Uncharacterized SufE-like protein YgdK	6.6
6	25.53	L-arabinose-binding periplasmic protein	6.6
6	28.53	Glyoxylate/hydroxypyruvate reductase A	6.8
6	15.7	Esterase FrsA	6.9
6	18.46	Chaperone SurA	7.0
6	19.37	Glucans biosynthesis protein G	7.2
6	78.26	Cold shock-like protein CspC	7.2
6	73.72	Uncharacterized protein YccU	7.3
6	34.51	N-acetylneuraminate epimerase	7.7
6	20.25	Periplasmic serine endoprotease DegP	8.6
5	19.18	Mannose-6-phosphate isomerase	5.4
5	19.8	Elongation factor Tu 1	5.5
5	27.4	Inner membrane protein yebE	5.5
5	44.44	Sugar phosphatase YfbT	6.1
5	48.87	Ribosome-binding factor A	6.2
5	35.16	Uncharacterized protein YcgM	6.3
5	26.19	Putative glucose-6-phosphate 1-epimerase	6.3
5	33.03	Glycine betaine-binding periplasmic protein	6.4
5	42.14	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	6.5
5	18.9	Arabinose 5-phosphate isomerase KdsD	6.7
5	19.19	Blue copper oxidase CueO	6.8
5	28.87	Universal stress protein D	6.9

5	21.01	Thiosulfate-binding protein	8.1
5	38.69	G/U mismatch-specific DNA glycosylase	9.1
5	18.26	Probable L,D-transpeptidase YnhG	9.4
5	48.94	50S ribosomal protein L25	9.6
5	36.02	Chaperone protein skp	9.7
4	12.04	60 kDa chaperonin	4.9
4	42.52	Autonomous glycyl radical cofactor	5.2
4	21.85	Sugar phosphatase YidA	5.3
4	22.39	Tryptophan synthase alpha chain	5.5
4	14.18	Succinyl-CoA ligase [ADP-forming] subunit beta	5.5
4	24.24	D-xylose-binding periplasmic protein	5.6
4	22.59	Transcriptional regulatory protein RstA	5.7
4	18.69	Cytoskeleton protein RodZ	5.8
4	32.69	Probable quinol monooxygenase YgiN	6.2
4	30.63	Sugar phosphatase YbiV	6.4
4	28.45	Transcriptional regulatory protein OmpR	6.4
4	23.53	Probable L,D-transpeptidase YbiS	6.4
4	24.42	3-hydroxydecanoyl-[acyl-carrier-protein] dehydratase	6.6
4	20.73	Uncharacterized protein YggE	6.6
4	15.05	Periplasmic AppA protein	6.7
4	33.5	Outer-membrane lipoprotein carrier protein	6.8
4	26.92	Acyl-CoA thioesterase I	7.6
4	10.35	Sulfite reductase [NADPH] hemoprotein beta-component	7.6
4	14.58	TPR repeat-containing protein YfgC	7.7
4	11.92	Periplasmic murein peptide-binding protein	8.4
4	47.83	Acylphosphatase	8.5
4	17.56	Uncharacterized protein YncE	9.2
4	19.62	50S ribosomal protein L3	9.9
4	35.42	50S ribosomal protein L15	11.2
3	11.11	Septum site-determining protein MinD	5.4
3	10.65	Enolase	5.5
3	14.24	Outer membrane protein assembly factor BamC	5.6
3	16.08	Quinone oxidoreductase 2	5.6
3	26.38	Nitrogen regulatory protein	5.9
3	15.43	Putative quinone oxidoreductase YhdH	5.9
3	12.09	Periplasmic pH-dependent serine endoprotease DegQ	5.9
3	45.71	Cold shock protein CspA	5.9
3	10.04	Protein yhjJ	6.0
3	17.7	Putative osmoprotectant uptake system substrate-binding protein osmF	6.1
3	11.38	Anhydro-N-acetylmuramic acid kinase	6.1
3	38.3	Galactitol-specific phosphotransferase enzyme IIB component	6.3
3	4.4	Protein yhgF	6.3
3	26.37	Probable Fe(2+)-trafficking protein	6.3

3	18.22	Glycerophosphoryl diester phosphodiesterase	6.5
3	6.45	Tail-specific protease	6.6
3	22.92	3-demethylubiquinone-9 3-methyltransferase	6.6
3	14.29	Cystine-binding periplasmic protein	6.6
3	18.4	Uncharacterized GST-like protein YghU	6.7
3	14.16	D-erythrose-4-phosphate dehydrogenase	6.8
3	19.03	Succinyl-CoA ligase [ADP-forming] subunit alpha	6.8
3	13.77	Cell division protein ZapD	6.8
3	9.97	D-allose-binding periplasmic protein	6.9
3	25.3	Periplasmic protein CpxP	6.9
3	25.69	Transcriptional regulator SlyA	7.1
3	22.05	Uncharacterized protein YhhK	7.1
3	50.6	Glutaredoxin-3	7.2
3	13.08	Riboflavin biosynthesis protein RibD	7.6
3	31.52	Chorismate--pyruvate lyase	8.0
3	11.91	Glutathione-binding protein GsiB	8.3
3	28.81	Protein YebF	8.6
3	22.63	Peptidyl-prolyl cis-trans isomerase A	8.9
3	40.13	Protein CreA	9.0
3	52.22	DNA-binding protein HU-alpha	9.6
3	22.33	30S ribosomal protein S10	9.7
3	39.73	Uncharacterized protein YhhA	10.9
2	36.51	Uncharacterized protein YbgS	4.9
2	45.61	RutC family protein YoaB	5.2
2	17.46	Uncharacterized metalloprotease YggG	6.1
2	21.09	DNA polymerase III subunit chi	6.2
2	4.46	Tryptophanase	6.2
2	12.5	Uncharacterized protein YeiR	6.5
2	13.6	Ribosomal RNA small subunit methyltransferase J	7.1
2	9.69	Right origin-binding protein	7.2
2	14.34	3-oxoacyl-[acyl-carrier-protein] reductase FabG	7.4
2	10.57	Uncharacterized protein YcbX	7.6
2	23.89	Uncharacterized protein YebY	8.0
2	5.78	Protein HemY	8.5
2	11.95	Sigma-E factor regulatory protein RseB	8.6
2	12.89	Peptidyl-tRNA hydrolase	9.0
2	40.37	UPF0265 protein YeeX	9.3
2	21.82	UPF0339 protein YegP	9.4
2	13.25	50S ribosomal protein L1	9.6
2	16.26	50S ribosomal protein L14	10.4

Table 3.3 $\Delta G^{o'}_{op}$ (kcal/mol) for the I29A;I37H variant in buffer (20 ^{o}C, pH 7.0, 50 mM sodium phosphate). Protons that exchange upon global unfolding (82) are bolded (SEM, standard error of the mean).

Residue	Trial 1	Trial 2	Trial 3	AVERAGE ±SEM
Trp5	6.56	6.79	6.72	6.69 ±0.07
Glu7	_	_	6.94	6.94
Leu8	6.37	6.58	6.57	6.51 ±0.07
Val9	5.88	6.11	6.03	6.01 ±0.07
Gly10	5.88	6.04	6.04	5.99 ±0.05
Lys11	**7.13**	**7.33**	**7.26**	**7.24 ±0.06**
Val13	6.11	6.30	6.27	6.23 ±0.06
Ala16	6.94	7.15	7.12	7.07 ±0.07
Lys17	_	_	5.26	5.26
Lys18	5.89	6.17	6.05	6.04 ±0.08
Val19	6.67	6.70	6.78	6.72 ±0.03
Ile20	**6.89**	**6.66**	**7.52**	**7.2 ±0.25**
Leu21	**6.38**	**6.30**	**6.72**	**6.47 ±0.13**
Gln22	6.63	6.84	6.79	6.75 ±0.06
Asp23	6.82	7.02	6.95	6.93 ±0.06
Lys24	6.51	6.69	6.67	6.62 ±0.06
Ala27	5.75	5.96	5.91	5.87 ±0.06
Ile30	**6.84**	**7.75**	**6.94**	**7.1 ±0.3**
Leu32	6.71	6.91	6.90	6.84 ±0.06
Val34	4.10	4.23	4.09	4.14 ±0.04
Arg46	6.72	6.94	6.89	6.85 ±0.07
Val47	**6.85**	**7.02**	**7.01**	**6.96 ±0.06**
Arg48	6.86	7.04	7.04	6.98 ±0.06
Leu49	**6.79**	**7.01**	**6.94**	**6.91 ±0.07**
Phe50	**6.75**	**6.79**	**6.84**	**6.79 ±0.03**
Val51	**6.58**	**6.81**	**6.66**	**6.68 ±0.07**
Asp52	6.57	6.83	6.72	6.71 ±0.08
Asp55	6.28	6.50	5.36	6.1 ±0.3
Asn56	7.38	7.60	7.54	7.51 ±0.07
Ala57	7.20	7.22	7.13	7.18 ±0.03
Ala58	6.77	7.01	6.87	6.88 ±0.07
Glu59	7.62	7.80	7.76	7.73 ±0.05
Arg62	6.71	6.90	6.88	6.83 ±0.06
Val63	6.80	6.93	6.96	6.90 ±0.05
Gly64	5.62	6.49	5.76	5.9 ±0.3

Table 3.4: $\Delta G^{o'}_{op}$ (kcal/mol) for the I29A;I37H variant in 100.0 g/L total-protein lysate (20 ^{o}C, pH 7.0, 50 mM sodium phosphate). Protons that exchange upon global unfolding (82) are bolded (SEM, standard error of the mean).

Residue	Trial 1	Trial 2	Trial 3	AVERAGE ±SEM
Trp5	6.07	5.98	5.94	6.00 ±0.04
Leu8	6.05	5.97	5.76	5.93 ±0.09
Val9	5.51	5.46	5.36	5.44 ±0.04
Lys11	**6.82**	**6.82**	**6.85**	**6.83 ±0.01**
Val13	5.72	5.75	5.73	5.73 ±0.01
Ala16	6.50	6.50	6.38	6.46 ±0.04
Lys18	5.30	5.65	5.46	5.5 ±0.1
Val19	6.14	6.10	5.94	6.06 ±0.06
Ile20	**6.32**	**6.19**	**5.84**	**6.1 ±0.14**
Leu21	**5.99**	**5.92**	**5.64**	**5.9 ±0.11**
Gln22	6.30	6.24	6.19	6.24 ±0.03
Asp23	6.51	6.52	6.54	6.52 ±0.01
Lys24	6.14	6.13	6.06	6.11 ±0.03
Ala27	5.44	5.37	5.41	5.41 ±0.02
Ile30	**6.25**	**6.21**	**5.97**	**6.14 ±0.09**
Leu32	6.32	6.27	6.07	6.22 ±0.08
Val34	_	4.53	5.08	4.8 ±0.6
Arg46	6.44	6.39	6.24	6.36 ±0.06
Val47	**6.46**	**6.36**	**6.23**	**6.35 ±0.07**
Arg48	6.61	6.50	6.47	6.53 ±0.04
Leu49	**6.49**	**6.47**	**6.34**	**6.43 ±0.05**
Phe50	**6.34**	**6.28**	**6.13**	**6.25 ±0.06**
Val51	**6.30**	**6.17**	**5.87**	**6.11 ±0.13**
Asp52	6.19	6.15	6.01	6.12 ±0.05
Ala57	6.54	6.50	6.33	6.46 ±0.06
Ala58	6.35	6.40	6.20	6.32 ±0.06
Glu59	7.10	7.06	6.92	7.03 ±0.05
Arg62	6.33	6.29	6.25	6.29 ±0.02
Val63	6.23	6.25	6.06	6.18 ±0.06
Gly64	5.16	5.10	4.76	5.0 ±0.12

Table 3.5: $\Delta G_{op}^{o'}$ (kcal/mol) for the I29A;I37H variant in 100.0 g/L anionic-protein lysate (20 ^{o}C, pH 7.0, 50 mM sodium phosphate). Protons that exchange upon global unfolding (82) are bolded (SEM, standard error of the mean).

Residue	Trial 1	Trial 2	Trial 3	AVERAGE ±SEM
Trp5	5.88	5.92	6.25	6.0 ±0.12
Leu8	5.93	5.68	6.20	5.9 ±0.15
Val9	5.47	5.37	5.74	5.5 ±0.1
Lys11	**6.87**	**6.72**	**7.05**	**6.9 ±0.1**
Val13	5.83	5.85	6.04	5.91 ±0.07
Ala16	6.72	6.44	6.73	6.63 ±0.1
Lys18	5.80	5.90	5.80	5.83 ±0.03
Val19	6.22	5.95	6.45	6.2 ±0.1
Ile20	**6.30**	**5.68**	**6.63**	**6.2 ±0.3**
Leu21	**5.87**	**5.65**	**6.23**	**5.9 ±0.17**
Gln22	6.23	6.29	6.59	6.4 ±0.11
Asp23	6.60	6.38	10.54	6.5 ±0.22
Lys24	6.28	6.22	6.43	6.31 ±0.06
Ala27	5.84	5.69	5.79	5.77 ±0.04
Ile30	**6.34**	**5.66**	**6.47**	**6.2 ±0.25**
Leu32	6.27	5.87	6.52	6.2 ±0.19
Val34	5.51	5.05	4.80	5.1 ±0.21
Arg46	6.49	6.36	6.69	6.5 ±0.1
Val47	**6.38**	**5.97**	**6.69**	**6.4 ±0.21**
Arg48	6.78	6.61	6.87	6.75 ±0.08
Leu49	**6.53**	**6.25**	**6.63**	**6.5 ±0.11**
Phe50	**6.48**	**6.25**	**6.61**	**6.4 ±0.11**
Val51	**6.19**	**5.86**	**6.45**	**6.2 ±0.17**
Asp52	6.29	6.09	6.41	6.26 ±0.09
Ala57	6.38	6.27	6.66	6.4 ±0.12
Ala58	6.35	6.14	6.65	6.4 ±0.15
Glu59	7.16	6.83	7.42	7.1 ±0.17
Arg62	6.47	6.29	6.57	6.44 ±0.08
Val63	6.17	6.11	6.48	6.2 ±0.11
Gly64	4.89	4.48	5.17	4.8 ±0.2

Table 3.6: Global stabilities of CI2.

	$\Delta G_{op}^{o'}$ (kcal/mol)	$\Delta\Delta G_{op}^{o'}$ (kcal/mol)
50 mM sodium phosphate, pH 6.5 or 7.0, 20 ^{o}C	6.9 ±0.1	_
100 g/L bovine serum albumin, pH 6.5, 20 ^{o}C	6.7 ±0.1	-0.2 ±0.1
100 g dry weight /L reconstituted lysate, pH 6.5, 20 ^{o}C	6.4 ±0.1	-0.5 ±0.1
100 g anionic proteins, pH 7.0, 20 ^{o}C	6.3 ±0.1	-0.5 ±0.1
100 g/L total proteins, pH 7.0, 20 ^{o}C	6.3 ±0.1	-0.6 ±0.1
130 g dry weight /L reconstituted lysate, pH 6.5, 20 ^{o}C	6.1 ±0.1	-0.8 ±0.1
50 mM sodium acetate, pH 5.4, 37 ^{o}C	4.9 ±0.1	_
100 g/L polyvinylpyrrolidone, pH 5.4, 37 ^{o}C	5.26 ±0.05	0.3 ±0.1
100 g/L Ficoll, pH 5.4, 37 ^{o}C	5.4 ±0.1	0.5 ±0.1

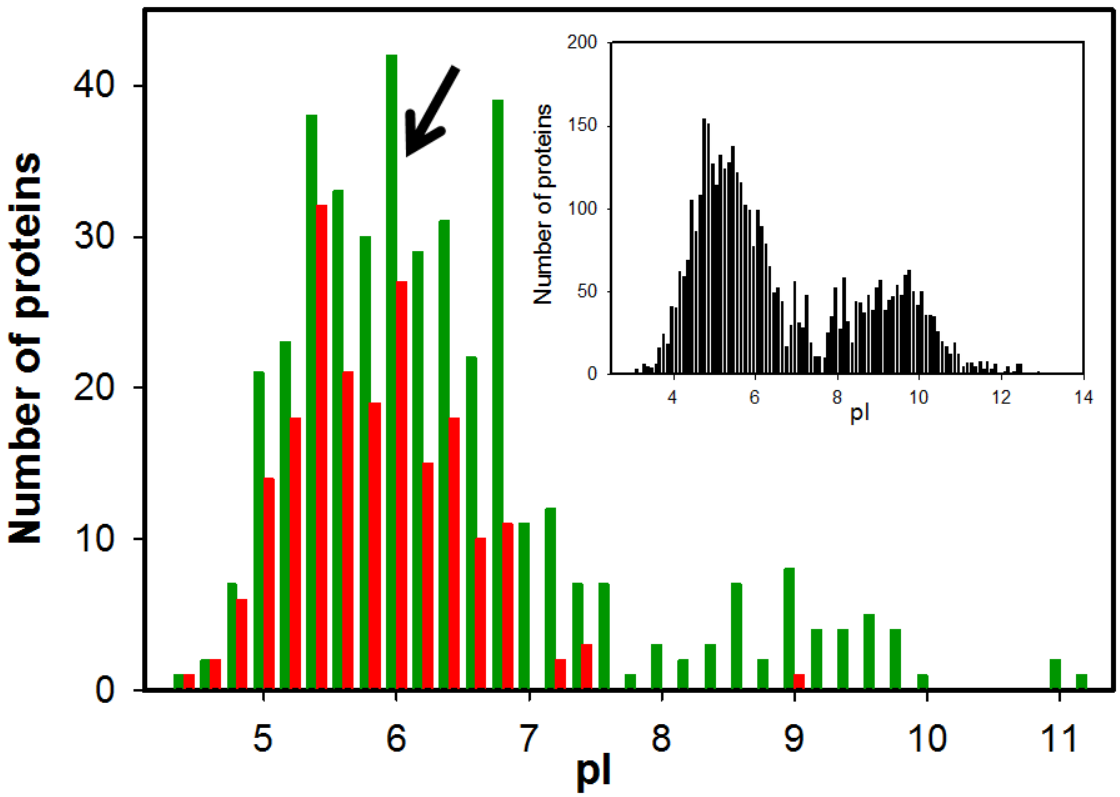

Figure 3.1 Histogram of proteins from the total protein lysate (green) and the anionic protein lysate (red) as a function of isoelectric point (pI). The arrow indicates the pI of CI2. The data are binned in 0.2 pI increments. The inset shows the distribution in the *E. coli* proteome (37).

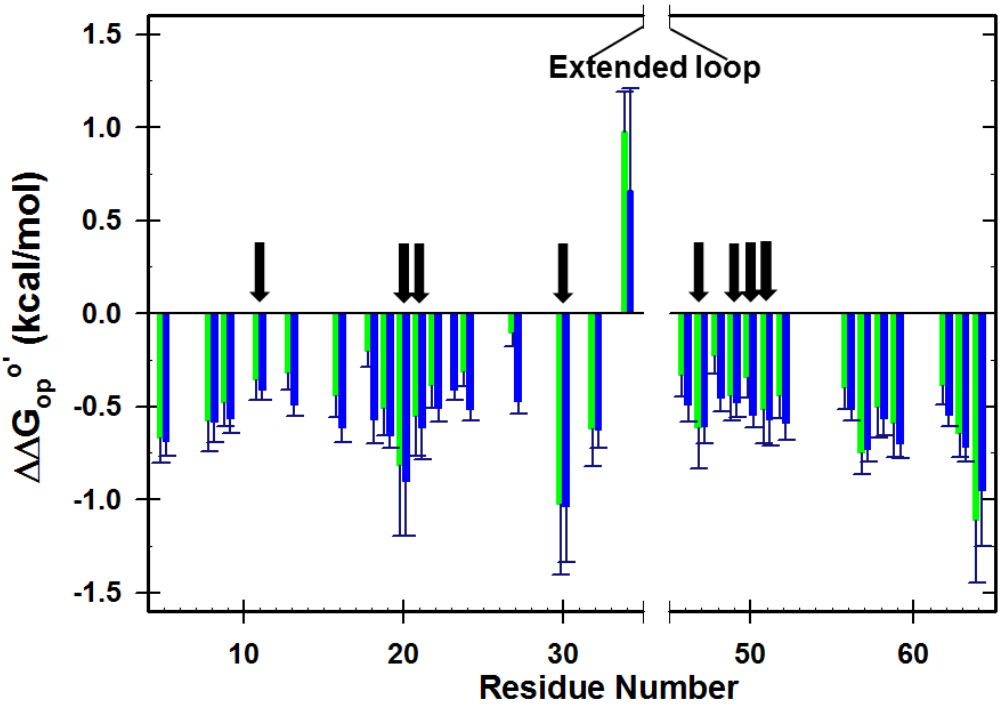

Figure 3.2: Bar graph of $\Delta\Delta G^{o'}_{op}$ (protein - buffer) for 100.0 g/L anionic-proteins (green) and 100.0 g/L total-proteins (blue) *versus* residue number (20 ^{o}C, pH 7.0, 50 mM sodium phosphate). Residues that exchange by global unfolding are indicated by black arrows. The values are the average of three trials at one condition minus the average of three trials at the other condition. The uncertainties are derived from error propagation on the standard deviations of the mean.

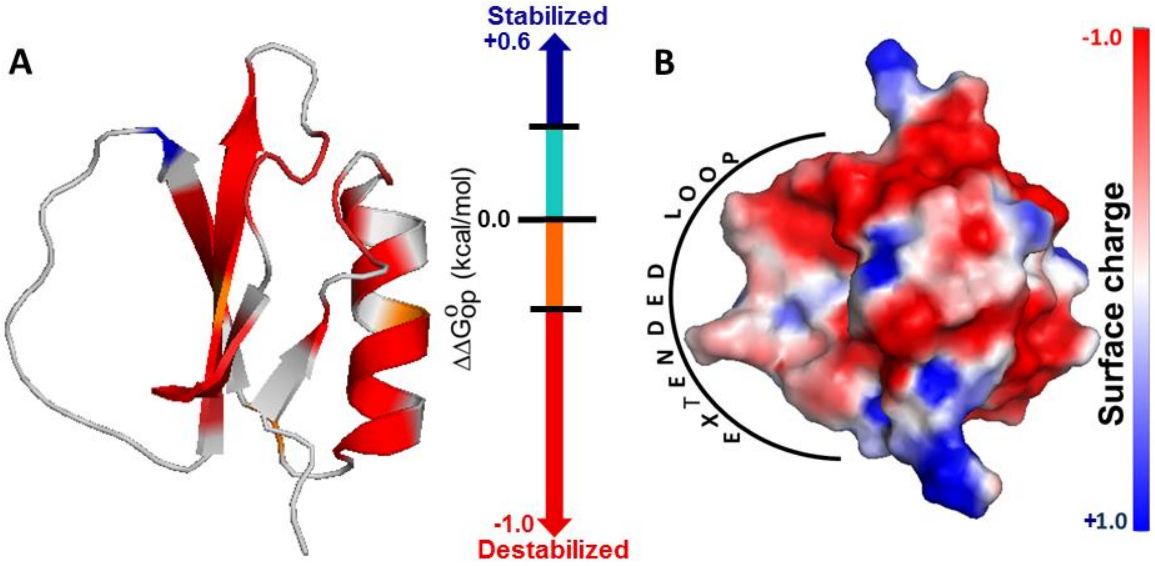

Figure 3.3: Structure of CI2(111) color coded by **(A)** $\Delta\Delta G^{o'}_{op}$ (lysate - buffer) in 100.0 g/L sample of anionic protein lysate and **(B)** electrostatic potential. The color-coded structure for $\Delta\Delta G^{o'}_{op}$ in 100.0 g/L total protein lysate is essentially identical to that shown. Charge was calculated using the Adaptive Poisson-Boltzmann Solver (120) in PyMOL (35), which approximates the potential felt by a point charge near the surface.

REFERENCES

1. Zimmerman SB & Trach SO (1991) Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of Escherichia coli. *Journal of Molecular Biology* 222(3):599-620.

2. Anfinsen CB (1973) Principles that govern the folding of protein chains. *Science* 181(96):223-30.

3. Richards FM (1977) Areas, volumes, packing, and protein structure. *Annual Review of Biophysics and Bioengineering* 6:151-76.

4. Fleming PJ & Rose JD (2008) *Conformational Properties of Unfolded Proteins* (Wiley-VCH).

5. Creighton TE (2010) *The Biophysical Chemistry of Nucleic Acids and Proteins* (Helvetian Press, 2010) p 774.

6. Minton AP (1981) Excluded volume as a determinant of macromolecular structure and reactivity. *Biopolymers* 20:2093-120.

7. Makhatadze GI & Privalov PL (1992) Protein interactions with urea and guanidinium chloride: a calorimetric study. *Journal of Molecular Biology* 226:491-505.

8. Timasheff SN (1993) The control of protein stability and association by weak interactions with water: how do solvents control these processes? *Annual Review of Biophysics and Biomolelcular Structure* 22:67-97.

9. Street TO, Bolen DW, & Rose GD (2006) A molecular mechanism for osmolyte-induced protein stability. *Proceedings of the National Academy of Sciences U.S.A.* 103(38):13997-4002.

10. Mayer JE (1942) Contribution to Statistical Mechanics. *Journal of Chemical Physics* 10(10):629-43.

11. Winzor DJ & Wills PR (1995) Thermodynamic non-ideality and protein interactions. *Protein-solvent Interactions*, ed Gregory RB (Marcel Dekker Inc, New York, U.S.A).

12. Lebowitz JL, Helfand E, & Praestgaard E (1965) Scaled Particle Theory of Fluid Mixtures. *Journal of Chemical Physics* 43:774-9.

13. Reiss H (1966) Scaled particle methods in the statistical thermodynamics of fluids. *Advances in Chemical Physics* 9:1-84.

14. Rubenstein M & Colby RH (2003) *Polymer Physics* (Oxford University Press Inc., New York).

15. Zhou HX, Rivas G, & Minton AP (2008) Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences. *Annual Review of Biophysics* 37:375-97.

16. Minton AP (2013) Quantitative assessment of the relative contributions of steric repulsion and chemical interactions to macromolecular crowding. *Biopolymers* 99(4):239-44.

17. Schellman J (2003) Protein stability in mixed solvents: a balance of contact interactions and excluded volume. *Biophysical Journal* 85:108-25.

18. Scatchard G (1946) Physical chemistry of protein solutions; derivation of the equations for the osmotic pressure. *Journal of American Chemical Society* 68(11):2315-9.

19. McMillan WG & Mayer JE (1945) The statistical thermodynamics of multicomponent systems. *Journal of Chemical Physics* 13:276-305.

20. Minton AP (1995) A molecular model for the dependence of the osmotic pressure of bovine serum albumin upon concentration and pH. *Biophysical Chemistry* 57(1):65-70.

21. Courtenay ES, Capp MW, Anderson CF, & Record MTJ (2000) Vapor pressure osmometry studies of osmolyte-protein interactions: implications for the action of osmoprotectants *in vivo* and for the interpretation of "osmotic stress" experiments *in vitro*. *Biochemistry* 39:4455-71.

22. Weatherly GT & Pielak GJ (2001) Second virial coefficients as a measure of protein--osmolyte interactions. *Protein Science* 10(1):12-6.

23. Davis-Searles PR, Saunders AJ, Erie DA, Winzor DJ, & Pielak GJ (2001) Interpreting the effects of small uncharged solutes on protein-folding equilibria. *Annual Review of Biophysics and Biomolecular Structure* 30:271-306.

24. Rivas G, Fernandez JA, & Minton AP (2001) Direct observation of the enhancement of noncooperative protein self-assembly by macromolecular crowding: indefinite linear self-association of bacterial cell division protein FtsZ. *Proceedings of the National Academy of Sciences U.S.A.* 98(6):3150-5.

25. Hirota S, *et al.* (2010) Cytochrome c polymerization by successive domain swapping at the C-terminal helix. *Proceedings of the National Academy of Sciences U.S.A.* 107(29):12854-9.

26. Jiao M, Li HT, Chen J, Minton AP, & Liang Y (2010) Attractive protein-polymer interactions markedly alter the effect of macromolecular crowding on protein association equilibria. *Biophysical Journal* 99(3):914-23.

27. Rule GS & Hitchens TK (2006) *Fundamentals of Protein NMR Spectroscopy* (Springer, Dorddrecht, The Netherlands).

28. Pielak GJ, *et al.* (2009) Protein nuclear magnetic resonance under physiological conditions. *Biochemistry* 48(2):226-34.

29. Barnes CO, Monteith WB, & Pielak GJ (2011) Internal and global protein motion assessed with a fusion construct and in-cell NMR spectroscopy. *ChemBioChem* 12(3):390-1.

30. Barnes CO & Pielak GJ (2011) In-cell protein NMR and protein leakage. *Proteins: Structure, Function, and Bioinformatics* 79(2):347-51.

31. Wang Q, Zhuravleva A, & Gierasch LM (2011) Exploring weak, transient protein-protein interactions in crowded *in vivo* environments by in-cell nuclear magnetic resonance spectroscopy. *Biochemistry* 50(43):9225-36.

32. Crowley PB, Chow E, & Papkovskaia T (2011) Protein interactions in the *Escherichia coli* cytosol: an impediment to in-cell NMR spectroscopy. *ChemBioChem* 12(7):1043-8.

33. Li C, *et al.* (2008) Differential dynamical effects of macromolecular crowding on an intrinsically disordered protein and a globular protein: implications for in-cell NMR spectroscopy. *Journal of American Chemical Society* 130(20):6310-1.

34. Li C & Pielak GJ (2009) Using NMR to distinguish viscosity effects from nonspecific protein binding under crowded conditions. *Journal of American Chemical Society* 131(4):1368-9.

35. Wang Y, Li C, & Pielak GJ (2010) Effects of proteins on protein diffusion. *Journal of American Chemical Society* 132(27):9392-7.

36. Dedmon MM, Patel CN, Young GB, & Pielak GJ (2002) FlgM gains structure in living cells. *Proceedings of the National Academy of Sciences U.S.A.* 99(20):12681-4.

37. Spitzer J & Poolman B (2009) The role of biomacromolecular crowding, ionic strength, and physicochemical gradients in the complexities of life's emergence. *Microbiology and Molecular Biology Reviews* 73(2):371-88.

38. Schlesinger AP, Wang Y, Tadeo X, Millet O, & Pielak GJ (2011) Macromolecular crowding fails to fold a globular protein in cells. *Journal of American Chemical Society* 133(21):8082-5.

39. Tadeo X, *et al.* (2009) Structural basis for the aminoacid composition of proteins from halophilic archea. *PLoS Biology* 7(12):e1000257.

40. Berg OG (1990) The influence of macromolecular crowding on thermodynamic activity: solubility and dimerization constants for spherical and dumbbell-shaped molecules in a hard-sphere mixture. *Biopolymers* 30(11-12):1027-37.

41. McGuffee SR & Elcock AH (2010) Diffusion, crowding & protein stability in a dynamic molecular model of the bacterial cytoplasm. *PLoS Computational Biology* 6(3):e1000694.

42. Charlton LM, *et al.* (2008) Residue-level interrogation of macromolecular crowding effects on protein stability. *Journal of American Chemical Society* 130(21):6826-30.

43. Miklos AC, Li C, Sharaf NG, & Pielak GJ (2010) Volume exclusion and soft interaction effects on protein stability under crowded conditions. *Biochemistry* 49(33):6984-91.

44. Sasahara K, McPhie P, & Minton AP (2003) Effect of dextran on protein stability and conformation attributed to macromolecular crowding. *Journal of Molecular Biology* 326(4):1227-37.

45. Stagg L, Zhang SQ, Cheung MS, & Wittung-Stafshede P (2007) Molecular crowding enhances native structure and stability of alpha/beta protein flavodoxin. *Proceedings of the National Academy of Sciences U.S.A.* 104(48):18976-81.

46. Christiansen A, Wang Q, Samiotakis A, Cheung MS, & Wittung-Stafshede P (2010) Factors defining effects of macromolecular crowding on protein stability: an in vitro/in silico case study using cytochrome c. *Biochemistry* 49(31):6519-30.

47. Benton LA, Smith AE, Young GB, & Pielak GJ (2012) Unexpected effects of macromolecular crowding on protein stability. *Biochemistry* 51(49):9773–5.

48. Crowley PB, Brett K, & Muldoon J (2008) NMR spectroscopy reveals cytochrome *c*-poly(ethylene glycol) interactions. *ChemBioChem* 9(5):685-8.

49. Zhang DL, Wu LJ, Chen J, & Liang Y (2012) Effects of macromolecular crowding on the structural stability of human alpha-lactalbumin. *Acta Biochimica et Biophysica Sinica* 44(8):703-11.

50. Pielak GJ & Miklos AC (2010) Crowding and function reunite. *Proceedings of the National Academy of Sciences U.S.A.* 107(41):17457-8.

51. Hermans J (1982) Excludedvolume theory of polymer-protein interactions based on polymer chain statistics. *Journal of Chemical Physics* 77:2193-203.

52. Homouz D, Perham M, Samiotakis A, Cheung MS, & Wittung-Stafshede P (2008) Crowded, cell-like environment induces shape changes in aspherical protein. *Proceedings of the National Academy of Sciences U.S.A.* 105(33):11754-9.

53. Miklos AC, Sarkar M, Wang Y, & Pielak GJ (2011) Protein crowding tunes protein stability. *Journal of American Chemical Society* 133(18):7116-20.

54. Feig M & Sugita Y (2012) Variable interactions between protein crowders and biomolecular solutes are important in understanding cellular crowding. *Journal of Physical Chemistry B* 116(1):599-605.

55. Harada R, Sugita Y, & Feig M (2012) Protein crowding affects hydration structure and dynamics. *Journal of American Chemical Society* 134(10):4842-9.

56. Lian L-Y (2013) NMR studies of weak protein-protein interactions. *Progress in Nuclear Magnetic Resonance Spectroscopy* 71:59-72.

57. Politi R & Harries D (2010) Enthalpically driven peptide stabilization by protective osmolytes. *Chemical Commununications* 46(35):6449-51.

58. Sukenik S, Sapir L, Gilman-Politi R, & Harries D (2013) Diversity in the mechanisms of cosolute action on biomolecular processes. *Faraday Discussions* 160:225-37.

59. Wang Y, Sarkar M, Smith AE, Krois AS, & Pielak GJ (2012) Macromolecular crowding and protein stability. *Journal of American Chemical Society* 134(40):16614-8.

60. Ghaemmaghami S & Oas TG (2001) Quantitative protein stability measurement in vivo. *Nature Structural Biology* 8(10):879-82.

61. Ignatova Z & Gierasch LM (2004) Monitoring protein stability and aggregation in vivo by real-time fluorescent labeling. *Proceedings of the National Academy of Sciences U.S.A.* 101(2):523-8.

62. Ignatova Z, *et al.* (2007) From the test tube to the cell: Exploring the folding and aggregation of a β-clam protein. *Biopolymers* 88(2):157-63.

63. Inomata K, *et al.* (2009) High-resolution multi-dimensional NMR spectroscopy of proteins in human cells. *Nature* 458(7234):106-9.

64. Ebbinghaus S, Dhar A, McDonald JD, & Gruebele M (2010) Protein folding stability and dynamics imaged in a living cell. *Nature Methods* 7(4):319-23.

65. Dhar A, *et al.* (2011) Protein stability and folding kinetics in the nucleus and endoplasmic reticulum of eucaryotic cells. *Biophysical Journal* 101(2):421-30.

66. Guo M, Xu Y, & Gruebele M (2012) Temperature dependence of protein folding kinetics in living cells. *Proceedings of the National Academy of Sciences U.S.A.* 109(44):17863-7.

67. Sarkar M, Li C, & Pielak GJ (2013) Soft interactions and crowding. *Biophysical Reviews* 5(2):187-94.

68. Knowles DB, LaCroix AS, Deines NF, Shkel I, & Record MT, Jr. (2011) Separation of preferential interaction and excluded volume effects on DNA duplex and hairpin stability. *Proceedings of the National Academy of Sciences U.S.A.* 108(31):12699-704.

69. Fodeke AA & Minton AP (2011) Quantitative characterization of temperature-independent and temperature-dependent protein-protein interactions in highly nonideal solutions. *Journal of Physical Chemistry B* 115(38):11261-8.

70. Zhou HX (2013) Influence of crowded cellular environments on protein folding, binding, and oligomerization: biological consequences and potentials of atomistic modeling. *FEBS Letters* 587(8):1053-61.

71. Phillip Y & Schreiber G (2013) Formation of protein complexes in crowded environments--from in vitro to in vivo. *FEBS Letters* 587(8):1046-52.

72. Wilf J & Minton AP (1981) Evidence for protein self-association induced by excluded volume. Myoglobin in the presence of globular proteins. *Biochimica et Biophysica Acta* 670(3):316-22.

73. Elcock AH (2010) Models of macromolecular crowding effects and the need for quantitative comparisons with experiment. *Current Opinion in Structural Biology* 20(2):196-206.

74. Neidhardt FC (1987) Chemical Composition of *Escherichia coli*. *Escherichia coli and Salmonella typhimurium,* (American Society of Microbiology, Washington, D. C.), Vol 1.

75. Blattner FR, *et al.* (1997) The complete genome sequence of *Escherichia coli* K-12. *Science* 277(5331):1453-62.

76. Gierasch LM & Gershenson A (2009) Post-reductionist protein science, or putting Humpty Dumpty back together again. *Nature Chemical Biology* 5(11):774-7.

77. Englander SW & Kallenbach NR (1983) Hydrogen exchange and structural dynamics of proteins and nucleic acids. *Quarterly Review of Biophysics* 16(4):521-655.

78. Miklos AC, Li C, & Pielak GJ (2009) Using NMR-detected backbone amide ^{1}H exchange to assess macromolecular crowding effects on globular-protein stability. *Methods in Enzymology* 466:1-18.

79. Smith AE, Sarkar M, Young GB, & Pielak GJ (2013) Amide proton exchange of a dynamic loop in cell extracts. *Protein Science 22*(10):1313-9.

80. Latham MP & Kay LE (2013) Probing non-specific interactions of Ca^{2}(+)-calmodulin in *E. coli* lysate. *Journal of Biomolecular NMR* 55(3):239-47.

81. Latham MP & Kay LE (2012) Is buffer a good proxy for a crowded cell-like environment? A comparative NMR study of calmodulin side-chain dynamics in buffer and E. coli lysate. *PLoS One* 7(10):e48226.

82. Itzhaki LS, Neira JL, & Fersht AR (1997) Hydrogen exchange in chymotrypsin inhibitor 2 probed by denaturants and temperature. *Journal of Molecular Biology* 270(1):89-98.

83. Link AJ, Robison K, & Church GM (1997) Comparing the predicted and observed properties of proteins encoded in the genome of *Escherichia coli* K-12. *Electrophoresis* 18(8):1259-313.

84. Charlton LM (2008) Protein behavior in crowded environments. Ph. D. (University of North Carolina at Chapel Hill, U.S.A).

85. Bai Y, Milne JS, Mayne L, & Englander SW (1993) Primary structure effects on peptide group hydrogen exchange. *Proteins* 17(1):75-86.

86. Nolting B, *et al.* (1997) The folding pathway of a protein at high resolution from microseconds to seconds. *Proceedings of the National Academy of Sciences U.S.A.* 94(3):826-30.

87. Wagner G (1980) A novel application of Nuclear Overhauser Enhancement (NOE) in proteins: analysis of correlated events in the exchange of internal labile protons. *Biochemical and Biophysical Research Communications* 97(2):614-20.

88. Fersht A (1999) *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding* (W.H. Freeman and Company, USA).

89. Itzhaki LS, Otzen DE, & Fersht AR (1995) The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding. *Journal of Molecular Biology* 254(2):260-88.

90. Chevelkov V, Xue Y, Rao DK, Forman-Kay JD, & Skrynnikov NR (2010) $^{15}N^{H/D}$-SOLEXSY experiment for accurate measurement of amide solvent exchange rates: application to denatured drkN SH3. *Journal of Biomolecular NMR* 46(3):227-44.

91. Hernandez G & LeMaster DM (2009) NMR analysis of native-state protein conformational flexibility by hydrogen exchange. *Methods in Molecular Biology* 490:285-310.

92. Skinner JJ, Lim WK, Bedard S, Black BE, & Englander SW (2012) Protein hydrogen exchange: testing current models. *Protein Science* 21(7):987-95.

93. Wagner G & Wüthrich K (1982) Amide protein exchange and surface conformation of the basic pancreatic trypsin inhibitor in solution. Studies with two-dimensional nuclear magnetic resonance. *Journal of Molecular Biology* 160(2):343-61.

94. Anderson JS, Hernandez G, & Lemaster DM (2008) A billion-fold range in acidity for the solvent-exposed amides of *Pyrococcus furiosus* rubredoxin. *Biochemistry* 47(23):6178-88.

95. Bennett BD, Yuan J, Kimball EH, & Rabinowitz JD (2008) Absolute quantitation of intracellular metabolite concentrations by an isotope ratio-based approach. *Nature Protocols* 3(8):1299-311.

96. Bennett BD, *et al.* (2009) Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nature Chemical Biology* 5(8):593-9.

97. Cayley S, Lewis BA, Guttman HJ, & Record MT, Jr. (1991) Characterization of the cytoplasm of *Escherichia coli* K-12 as a function of external osmolarity. Implications for protein-DNA interactions in vivo. *Journal of Molecular Biology* 222(2):281-300.

98. Burg MB & Ferraris JD (2008) Intracellular organic osmolytes: function and regulation. *Journal of Biological Chemistry* 283(12):7309-13.

99. Zhou HX (2013) Polymer crowders and protein crowders act similarly on protein folding stability. *FEBS Letters* 587(5):394-7.

100. McConkey EH (1982) Molecular evolution, intracellular organization, and the quinary structure of proteins. *Proceedings of National Academy of Sciences U.S.A.* 79(10):3236-40.

101. Srere PA (2000) Macromolecular interactions: tracing the roots. *Trends in Biochemical Sciences* 25(3):150-3.

102. Greenfield D, *et al.* (2009) Self-organization of the *Escherichia coli* chemotaxis network imaged with super-resolution light microscopy. *PLoS Biology* 7(6):e1000137.

103. Spitzer J & Poolman B (2013) How crowded is the prokaryotic cytoplasm? *FEBS Letters*.

104. Wittekind M & Mueller L (1993) HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the alpha-carbon and beta-carbon resonances in proteins. *Journal of Magnetic Resonance Series B* 101(2):201-5.

105. Grzesiek S & Bax A (1992) Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *Journal of American Chemical Society* 114:6291-3.

106. Delaglio F, *et al.* (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR* 6(3):277-93.

107. Johnson BA & Blevins RA (1994) NMR View: A computer program for the visualization and analysis of NMR data. *Journal of Biomolecular NMR* 4(5):603-14.

108. Schowen KB & Schowen RL (1982) Solvent isotope effects of enzyme systems. *Methods in Enzymology* 87:551-606.

109. Consortium TU (2012) Reorganizing the protein space at the Universal Protein Resource (UniProt). in *Nucleic Acids Research*, pp D71-D5.

110. Davison TS, *et al.* (2001) Structure and functionality of a designed p53 dimer. *Journal of Molecular Biology* 307(2):605-17.

111. McPhalen CA & James MN (1987) Crystal and molecular structure of the serine proteinase inhibitor CI-2 from barley seeds. *Biochemistry* 26(1):261-9.

112. Cavallo L, Kleinjung J, & Fraternali F (2003) POPS: A fast algorithm for solvent accessible surface areas at atomic and residue level. *Nucleic Acids Research* 31(13):3364-6.

113. Sarkar M, Smith AE, & Pielak GJ (2013) Impact of reconstituted cytosol on protein stability. *Proceedings of National Academy of Sciences U.S.A.* 110(48):19342-7.

114. Li C, Wang Y, & Pielak GJ (2009) Translational and rotational diffusion of a small globular protein under crowded conditions. *Journal of Physical Chemistry B* 113(40):13390-2.

115. Liang J, *et al.* (2009) Effective elimination of nucleic acids from bacterial protein samples for optimized blue native polyacrylamide gel electrophoresis. *Electrophoresis* 30(14):2454-9.

116. Taylor JR (1982) *An Introduction to Error Analysis: The Study of Uncertainities in Physical Measurements* (University Science Books, Sausalito, CA).

117. Kiraga J, *et al.* (2007) The relationships between the isoelectric point and: length of proteins, taxonomy and ecology of organisms. *BMC Genomics* 8:163.

118. Roesler KR & Rao AG (1999) Conformation and stability of barley chymotrypsin inhibitor-2 (CI-2) mutants containing multiple lysine substitutions. *Protein Engineering* 12(11):967-73.

119. Zhang Y-Z (1995) Protein and peptide structure and interactions studied by hydrogen exchange and NMR. Ph. D. (University of Pennsylvania, U.S.A).

120. Baker NA, Sept D, Joseph S, Holst MJ, & McCammon JA (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences U.S.A.* 98(18):10037-41.